Department of Obstetrics and Gynecology
Helsinki University Central Hospital

and

Doctoral Program of Clinical Research
Faculty of Medicine
University of Helsinki
Finland

QUALITY AND PATIENT SAFETY IN OBSTETRIC CARE

BENCHMARKS FOR IMPROVEMENT

Aura Pyykönen

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of
the University of Helsinki, for public examination in the Seth Wichmann Auditorium,
Department of Obstetrics and Gynecology, Helsinki University Central Hospital,
Haartmaninkatu 2, Helsinki, on Friday September 8th 2017, at 12 noon.

Helsinki 2017

Supervised by

Professor Lasse Lehtonen, MD, PhD
Department of Public Health
University of Helsinki and Helsinki University Hospital, Finland

Docent Anna-Maija Tapper, MD, PhD
Hyvinkää Hospital
University of Helsinki and Helsinki University Hospital, Finland

Reviewed by

Docent Jukka Uotila, MD, PhD
Department of Obstetrics and Gynecology
Tampere University Hospital, Finland

Docent Kaarin Mäkikallio, MD, PhD
Department of Obstetrics and Gynecology
Turku University Hospital, Finland

Official Opponent

Associate Professor Jens Langhoff-Roos, MD, PhD
Department of Obstetrics
Rigshospitalet, Copenhagen, Denmark

ISBN 978-951-51-3561-2 (pbk.)

ISBN 978-951-51-3562-9 (PDF)

Unigrafia

Helsinki 2017

To Lauri-Kustaa, Alma, and Amos

ABSTRACT

Optimal care of pregnant women and newborn babies has an immeasurable effect on the well-being of a family, but it also significantly impacts society overall. In Finland, the quality of obstetric and perinatal care is high, but in order to further improve it, sensitive and commonly agreed quality indicators are required.

The aim of the thesis was to validate different potential quality indicators for obstetric care, and to assess differences between different hospital size-categories in Finland. To facilitate reliable comparisons for benchmarking, we evaluated the usability of the Robson classification, a system for grouping the parturient population. This was done as an international comparison between the Nordic countries. The thesis also aims to contribute to more unified and high-quality obstetric care by examining the optimal time of labor induction in prolonged pregnancies. This thesis is based on data from the medical birth registries in Finland and in the other Nordic countries.

The findings of this thesis indicate true differences in the obstetric management and treatment culture between the Nordic countries and between different-sized birth units in Finland. However, it should be acknowledged that part of the differences may be explained by confounding background factors. The results enforce the existing evidence on the effect of birth centralization on neonatal outcomes, e.g. lower neonatal mortality in large units, but indicate a higher risk for instrumental delivery. Obstetric trauma, a potentially preventable complication of vaginal delivery, showed substantial size-dependent variation between the Finnish birth units; the risk was the lowest in mid-sized birth units. The Robson classification proved to bed usable in the Nordic setting, facilitating more accurate comparisons, but comprehensive interpretations require experience and expertise.

In our study on the effect of labor induction in prolonged pregnancy, we found an increased risk for Cesarean section around 41 gestational weeks, but no longer when the gestational age approached 42 weeks. From a neonatal perspective, labor induction resulted in a decreased risk for meconium aspiration syndrome, but only when the intervention took place before 41^{+5} weeks. Labor induction did not have a significant effect on neonatal mortality.

International and inter-unit benchmarking pinpoints the areas needing improvement; e.g. prevention of obstetric trauma or induction protocols among nulliparous women. Some indicators, such as Cesarean section rates within Robson groups, show potential for readily directing obstetric practices, and are therefore of special interest. Like our neighboring countries Denmark and Sweden, we should establish a national quality monitoring program for obstetric care in Finland, too, to ensure high-quality care also in the future.

FINNISH SUMMARY

Onnistunut raskaana olevien ja vastasyntyneiden hoito vaikuttaa merkittävästi yksilön ja perheen hyvinvointiin, ja samalla sillä on kauaskantoisia vaikutuksia koko yhteiskuntaan. Nyky-Suomessa synnytysten hoito on korkealaatuista, mutta hoidon kehittämisen kannalta on oleellista, että käytössä on riittävän herkkiä, yhteisesti sovittuja laatumittareita.

Tässä väitöstutkimuksessa pyrimme testaamaan käyttökelpoisia mittareita havainnollistamalla eroja Suomen synnytyssairaaloiden välillä kokoluokittain. Vertailukelpoisuuden tukemiseksi tutkimme pohjoismaisia keisarileikkauslukuja Robson-järjestelmän avulla: Järjestelmä mahdollistaa synnyttäjien luokittelun ja lopputulosten vertailun yhdenmukaistetuissa väestöissä. Pyrimme myös tukemaan yhdenmukaista ja korkealaatuista hoitoa selvittämällä yliaikaisten raskauksien optimaalista synnytyksenkäynnistysajankohtaa. Väitöskirjan osatyöt perustuvat Suomen ja muiden Pohjoismaiden syntymärekisteritietoihin.

Tutkimuksessa havaittiin merkittäviä eroja äidin ja lapsen muuttujissa sekä kansainvälisesti Pohjoismaissa, että sairaalakokoluokkien välillä Suomessa. Osa eroista selittyy taustamuuttujilla, mutta löydökset viittaavat todellisiin eroihin toimintakulttuurissa. Tuloksemme Suomesta tukevat aikaisempia löydöksiä keskittämisen eduista vastasyntyneen kannalta, mikä näkyi ennen kaikkea suurten yksiköiden matalampana neonataalikuolleisuutena. Toisaalta, tulokset viittaavat suurten yksiköiden korkeampaan toimenpiteellisten synnytysten riskiin. Sulkijalihasrepeämien, joita pidetään pääosin estettävissä olevana alatiesynnytysten komplikaationa, osuus vaihteli sairaalakokoluokittain siten, että riski oli pienin keskikokoisissa sairaaloissa. Robson-luokitus tuki kansainvälistä vertailua, mutta tulosten kokonaisvaltainen tulkinta osoittautui haastavaksi ja järjestelmän laajamittainen käyttö vaatii asiantuntemusta.

Yliaikaisessa raskauksien osalta tuloksemme osoittivat, että käynnistys lisää keisarileikkausriskiä 41 raskausviikon tuntumassa, mutta ei enää lähellä 42 raskausviikkoa. Käynnistys pienensi vastasyntyneiden meconium-aspiraatioriskiä ennen 41^{+5} raskausviikkoa, mutta vaikutusta ei ollut havaittavissa enää sen jälkeen. Kuolleisuuteen käynnistyksellä ei ollut vaikutusta.

Yksiköiden välinen vertailu (benchmarking) laatumittareiden avulla, kuten sulkijalihasrepeämien esiintyvyys tai ensisynnyttäjien käynnistysprosentti, tuo esiin niitä osa-alueita, joita on tarpeen kehittää. Laadun seurannassa painotus tulisi olla nimenomaan toimintaa ohjaavissa mittareissa, joista yhtenä esimerkkinä on keisarileikkausprosenttien tarkastelu Robson-ryhmien avulla. Jotta perinataalihoito säilyisi Suomessa korkealaatuisena myös tulevaisuudessa, tulisi meidänkin Tanskan ja Ruotsin tavoin luoda kansallinen laadunseurantaohjelma.

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

I Pyykönen A, Gissler M, Løkkegaard E, Bergholt T, Rasmussen SC, Smárason A, Bjarnadóttir RI, Másdóttir BB, Källén K, Klungsoyr K, Albrechtsen S, Skjeldestad FE, Tapper AM. Cesarean section trends in the Nordic Countries - a comparative analysis with the Robson classification. Acta Obstet Gynecol Scand. 2017 May;96(5):607-616.

II Pyykönen A, Gissler M, Jakobsson M, Lehtonen L, Tapper AM. The rate of obstetric anal sphincter injuries in Finnish obstetric units as a patient safety indicator. Eur J Obstet Gynecol Reprod Biol. 2013 Jul;169(1):33-8.

III Pyykönen A, Gissler M, Jakobsson M, Petäjä J, Tapper AM. Determining obstetric patient safety indicators: the differences in neonatal outcome measures between different-sized delivery units. BJOG: An International Journal of Obstetrics & Gynaecology. 2014 Mar;121(4):430-7.

IV Pyykönen A, Tapper AM, Gissler M, Haukka J, Petäjä J, Lehtonen L. The optimum time of labor induction for prolonged pregnancies - assessing the outcomes with propensity score method. Accepted for publication in Acta Obstet Gynecol Scand on 15 Aug, 2017.

The publications are referred to in the text by their roman numerals. They have been reprinted with the permission of their copyright holders. In addition, some unpublished material is presented.

ABBREVIATIONS

ACOG	American College of Obstetricians and Gynecologists
AHRQ	Agency for Healthcare Research and Quality
AOI	Adverse Outcome Index
BMI	Body Mass Index
CS	Cesarean section
DNIP	Danish National Indicator Project
EXPH	Expert panel on effective ways of investing in health
FET	Frozen embryo transfer
GA	Gestational age
ICD-10	International Classification of Diseases, 10th revision
ICSI	Intracytoplasmic sperm injection
IOM	Institute of Medicine
IVF	In vitro fertilization
MBR	Medical Birth Register
NAOI	Neonatal Adverse Outcome Indicator
NICU	Neonatal Intensive Care Unit
OBSQID	Obstetric quality project by the European Union
OECD	Organisation for Economic Co-operation and Development
OR	Odd Ratio
PS	Propensity Score
QI	Quality improvement
RCOG	Royal College of Obstetricians and Gynaecologists
RR	Relative Risk
SD	Standard deviation
TGCS	The ten-group classification system (also known as the Robson classification system)
THL	National Institute of Health and Welfare
uc-pH	Umbilical cord pH
UN	United Nations
VAS	Visual Analog Scale
VBAC	Vaginal birth after a previous Cesarean section
VTT	Technical Research Center of Finland
WAOS	Weighted Adverse Outcome Index
WHO	World Health Organization

DEFINITIONS

These terms, which are ambiguous and imprecise, are used in this thesis to mean the following:

Birth trauma: injury to neonate resulting from mechanical forces during birth. As a patient safety indicator, birth trauma comprehends several independent traumatic conditions, e.g. brachial plexus injury and cephalhematoma (1, 2)

Late-term: gestational age between 41 and 42 weeks (41^{+0} – 41^{+6})

Obstetric trauma: third- and fourth-degree perineal lacerations (involving trauma to anal sphincter)

Obstetric volume of a birth unit: number of annual deliveries

Postterm: gestational age ≥ 42^{+0}

Pre-labor Cesarean section: CS before the onset of labor (elective CS)

Prolonged hospitalization of neonate: newborn hospitalized ≥ 7 days after birth

Prolonged pregnancy: gestational age ≥40^{+0}

Term delivery: gestational age ≥ 37^{+0}

INTRODUCTION

No goals in healthcare are more critical than keeping patients safe from harm and improving the delivery and outcomes of their care (3). Providing the best-known care, evidence-based medicine, is the cornerstone of medical practice. High-quality healthcare, however, goes beyond this. Apart from being effective (evidence-based medicine), the care must be safe, patient-centered, timely, efficient, and equitable (4). Understanding this multidimensional nature of quality is crucial for assessing and improving quality of care.

Obstetric and perinatal health reflect the health of a society and are therefore of special interest for healthcare quality assessment. Lee Jong-wook, the former director general of the World Health Organization (WHO) stated: *"Mothers, the newborn and children represent the well-being of a society and potential for the future. Their health needs cannot be left unmet without harming the whole of society."(5)* It is justified to state that quality improvement in obstetric and perinatal care will have a significant, long-term impact on society.

Quality chasm, the gap between evidence and practice i.e. what we know and what we deliver, exists in obstetrics and in perinatal care (6). The use of different treatment practices varies broadly across facilities, providers, and geographic areas. This is primarily due to differences in treatment culture and other extrinsic factors rather than differences in needs of mothers and newborns (7). Lack of consensus on the best practice protocols is one of the main challenges in achieving standardized, evidence-based obstetric care (8).

Traditionally, obstetric care has relied on innovation rather than evidence (9, 10). In 1979, Archie Cochrane judged obstetrics to be the specialty making the worst use of randomized controlled trials (although he later acknowledged vast improvements in the specialty) (11). Many obstetric practices, such as episiotomy (12) and electronic fetal monitoring (13), have not only been adopted without prior evaluation, but are used continuously despite conflicting empirical and scientific evidence on the effects. When there is existing data on best practice, the implementation may prove difficult, causing underuse or delayed adoption of beneficial treatment protocols (8), e.g. corticosteroids in preterm delivery, or conversely, the beneficial treatment protocols may be overused (7, 14).

Over the past decades, centralization of births, both high- and low-risk, has indisputably improved neonatal outcomes (15-18). This development merits appraisal. Apart from an immediate increase in the well-being of a family, a decrease in neonatal morbidity results in substantial, long-term benefits for society as a whole. From another perspective, the evolution to

fewer but larger birth units may increase the risks due to unplanned out-of-hospital deliveries (19), and the impact of centralization on the quality of low-risk maternal care remains poorly studied (20). Medicalization of birth seems immediately connected to the expanding obstetric volumes and is accompanied by increased operative delivery rates and decreased rates of births without intervention (21).

Assessment of quality requires standardized quality indicators. Due to the dual nature of obstetrics and perinatal care, the fact that maternal and neonatal needs are often not aligned, it is imperative to measure the quality from both perspectives. Traditional examples are maternal and perinatal mortality, often seen as representations of the accessibility and effectiveness of a country's health system. Over the past 15 years, with increasing concerns about healthcare quality, more sensitive outcome indicators have been introduced, e.g. obstetric trauma (third- and fourth-degree perineal lacerations) and birth trauma on neonate (1, 22). The outcome indicators should be complemented by robust process measures, but their development has been hindered by the complex issue of defining the best practice (8). Recent examples of process measures as quality indicators are Cesarean section (CS) rates on different stratified, risk-adjusted populations and establishment of skin-to-skin contact between the mother and the neonate (23, 24).

It is reassuring that amidst the global increase in birth interventions combined with no or very little improvement in outcomes, the Finnish figures have good standing in international comparisons. Obstetric trauma rates in Finland are among the lowest in Europe (25), as are CS rates (26) and combined proportions of operative deliveries (25). Following an international trend, but based on conflicting evidence of labor inductions effect on outcomes (27, 28), the rate of induced labors has rapidly increased in Finland, from 20% in 2010 to 25% in 2015 (29).

There is a long tradition of high-quality medical birth register (MBR) and outcome reporting on a national level in Finland (30, 31). The other Nordic countries have similar MBRs, offering an extensive amount of data for benchmarking. These data are, however, underused for quality improvement purposes. The reported differences in outcome rates highlight areas ripe for development and could be used for adjusting guidelines and best practice protocols. This thesis aims to enhance obstetric quality management in Finland by examining international as well as inter-unit differences using both recognized and promising new quality indicators.

REVIEW OF THE LITERATURE

1 CONCEPT OF HEALTHCARE QUALITY AND PATIENT SAFETY

1.1 DEFINITION OF HEALTHCARE QUALITY

The concept of quality is essentially an evaluation of whether a product is good enough and best suited for its purpose. Measurement, benchmarking, and improvement – aimed at the highest achievable standard of excellence – are imperative parts of the concept of quality.

There are various published definitions for quality of healthcare (32). The most influential has likely been the one provided by the Institute of Medicine (IOM) in 1990: "…the degree to which health care services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge"(33). The definition incorporates the idea of the multidimensional nature of healthcare quality, which IOM defined more clearly a decade later (see below) (4). The definition encompasses a wider range of elements than the actual medical care given by referring to "healthcare services". By "desired health outcome" and "current professional knowledge" it refers to the prevailing idea of healthcare standards, both in the form of effectiveness of care and in what is known to be appropriate or best care.

The definition by the Council of Europe from 1998 emphasizes one dimension of quality, safety of care, more strongly than IOM: "Quality of care is the degree to which the treatment dispensed increases the patient's chances of achieving the desired results and diminishes the chances of undesirable results (safety), having regard to the current state of knowledge."(34). The World Health Organization (WHO)has not provided only a single definition for healthcare quality but defines quality by its dimensions (35).

1.2 MULTIDIMENSIONAL NATURE OF HEALTHCARE QUALITY

The six dimensions introduced by the IOM (4) are likely to be the most widely recognized determinants for healthcare quality (Table 1). Table 1 presents two other perspectives for quality dimensions parallel to the IOM's

definition (35, 36). As shown, the dimensions are largely the same but each dimension is emphasized slightly differently.

Safety is an imperative part of quality but also a discipline of its own and often assessed independently (see below). Effectiveness is the other key dimension to which both the outcome-oriented traditions in the assessment of medical care and the abundance of suitable measurement tools have contributed (37). Effectiveness comprehends also the essential principle of evidence-based medicine; we need to provide exclusively care that it is shown to be beneficial.

1.3 PATIENT SAFETY

Quality and patient safety are often discussed together. They are interdependent, both rooted in the same period in history, but in modern quality thinking, quality is regarded as a broader concept and includes patient safety as one of its imperative dimensions (4, 38, 39). IOM defines patient safety comprehensively, but somewhat imprecisely as "the prevention of harm to patients" (40). Emanuel et al (38) have provided a more detailed definition "Patient safety is a discipline in the healthcare sector that applies safety science methods toward the goal of achieving a trustworthy system of healthcare delivery. Patient safety is also an attribute of healthcare systems; it minimizes the incidence and impact of, and maximizes recovery from, adverse events."

An adverse event, defined by IOM, "results in unintended harm to the patient by an act of commission or omission rather than by the underlying disease or condition of the patient." A near miss, in turn, "is an act of commission or omission that could have harmed the patient but did not cause harm as a result of chance, prevention, or mitigation" (40).

A defining moment in the development of the modern concept of patient safety was the understanding that errors do occur despite all the known power of modern medicine. This was the key message in the groundbreaking IOM publication *To Err is Human* in 1999 (41). In the publication, IOM claimed that in the United States, there were close to 100,000 annual deaths due to preventable medical errors but, importantly, most of these errors were not a result of incompetent healthcare professionals but rather a result of poor systems and processes that fail to prevent errors (see *Systems thinking* below).

Table 1 *Dimensions of quality according to the Institute of Medicine (IOM), The World Health Organization (WHO) and the Expert panel on effective ways of investing in health (EXPH) for the European Commission*

IOM: High-quality healthcare is	WHO: High-quality healthcare system is	EXPH: All health services should be
1. Safe, avoiding harm to patients from the care that is intended to help them.	**6. Safe**, delivering healthcare that minimizes risks and harm to service users.	**2. Safe**, preventing avoidable harm related to care
2. Effective, providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and misuse, respectively).	**1. Effective**, delivering healthcare that is adherent to an evidence-based medicine and results in improved health outcomes for individuals and communities, based on need	**1. Effective**, and improve health outcomes
3. Patient-centered, providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions	**4. Acceptable/patient-centered**, delivering healthcare that takes into account the preferences and aspirations of individual service users and the cultures of their communities	**4. Patient-centered**, and involve patients/people as key partners in the process of care
4. Timely, reducing waits and sometimes harmful delays for both those who receive and those who give care.	**3. Accessible**, delivering healthcare that is timely, geographically reasonable, and provided in a setting where skills and resources are appropriate to medical need	**3. Appropriate**, and comply with current professional knowledge as well as meeting agreed standards
5. Efficient, avoiding waste, including waste of equipment, supplies, ideas, and energy.	**2. Efficient**, delivering healthcare in a manner which maximizes resource use and avoids waste	**5. Efficient and equitable**, and lead to the best value for the money spent and to equal access to available care for equal need, utilization and equal quality of care for all
6. Equitable, providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status	**5. Equitable**, delivering healthcare which does not vary in quality because of personal characteristics such as gender, race, ethnicity, geographical location, or socioeconomic status	

1.3.1 From innovation to evidence based quality chasm – History of healthcare quality

A public health pioneer, Florence Nightingale, marked the first step in the documented history of healthcare quality improvement. She suggested that the lack of sanitation and hygiene standards would be the cause of high mortality among wounded soldiers during the Crimean war in the mid-1800s. Her innovations were simple but groundbreaking: she promoted hand washing and the use of sanitized surgical tools, good nutrition and fresh air. Most importantly, she worked systematically and kept meticulous records, and indeed, her system of record keeping and tabulation formed a basis for the statistical quality measurement of today (42).

During most of the following century it was improvisation and innovation, e.g. penicillin, essential vaccines and x-ray, not systematic development, that lead to great improvements in the quality of medical care. Industrial engineering advanced quality thinking in the mid-1900s with ideas that were later adapted to healthcare (42). The physicists Walter A. Sewhart and W. Edwards Deming (43) and engineer Joseph M. Juran (44) developed concepts of quality planning, quality control, and quality improvement with tools like standardized work processes and data-driven decision-making, which seem obvious today but were revolutionary at the time (45).

In 1966, A. Donabedian, influenced by the ideas of Sewhart, Deming, and Juran, published the iconic paper "*Evaluating the Quality of Medical Care*" (46): There, he introduced the classic framework for measuring quality in healthcare according to three dimensions: 1) structure – the framework of care, 2) process – the actual care given, and 3) outcome of care. This has served as the basis for quality measurement over the past decades (47-50) – and is still used, also for evaluating obstetric care settings (24, 51).

Along with Donabedian, considered the founder of the study of healthcare quality, A. Cochrane contributed significantly to quality as well, his focus being on effective, evidence-based care (52). The principles he set forth 40 years ago are still valid today; he stated that resources will always be limited, and therefore, healthcare should aim to provide care equitably and only in the forms reliably shown to be effective. He emphasized the importance of making the best use of knowledge gained by the modern randomized controlled trials (RCT), the first of which was published in 1946, only a couple of decades earlier (53).

After 2000, the study of healthcare quality has expanded to encompass nearly all medical specialties. The single most influential factor for this development is likely the IOM publication "Crossing the quality chasm: A new health system for the 21st century" in 2001 (4). The publication claimed there is not only a gap, but a chasm between what we know and what we deliver. WHO rephrased the same concern in 2006, stating that the expected

health outcomes are not predictably achieved even in the highly developed and resourced healthcare systems (35). At the core of these publications was in healthcare systems had not responded to the rapid increases in medical knowledge, and therefore, delivered suboptimal care. Today, emphasis of quality has shifted further to value-based assessment: "more with less" (4, 32, 54, 55).

1.4 MEASURING QUALITY

1.4.1 Selecting quality indicators

There is an exhaustive number of papers addressing the prerequisites of quality indicators for healthcare. A common factor is the strong emphasis placed on the *validity of the indicator,* which can be further divided into the following categories:

- *Construct validity*: Theory or tradition supports that the indicator effectively measures what it claims to measure (50, 56-61)
- *Content validity:* Indicator is evidence-based and measures all aspects of what it intends to measure (50, 56, 59, 60)
- *Face validity***:** Both the physicians and the patients believe that the indicator measures quality of care (50, 57-59)

Other important qualities of recommendable quality indicators:

- linked to a national (or local) goal: increases the face-validity of the indicator as well as user engagement (59)
- clearly defined (50, 56)
- variation in outcome rates (50, 60)
- easily collected: Ideal measure is drawn from existing or easily collected data and the data collection does not impose undue burden on care providers (50, 60)

1.4.2 Dimensions of quality measurement – Donabedian model

In the Donabedian model, healthcare is measured in three dimensions (46):

- Structure; the framework of care
- Process; the actual care given
- Outcomes; the consequence of care

Process and structure form the care given. They are where the quality improvement takes place and that is why these elements should be in the core of quality measurement. The outcomes, in turn, may or may not reflect the effect of care, but as Donabedian stated "…(outcomes) remain the

ultimate validators of the effectiveness and quality of medical care" and today's quality assessment widely relies on outcome measures (46).

Outcome measures are easy to define and collect, they are conceptually straightforward, and tend to have high face validity (62). Their main limitation is a possible lack of construct validity; outcomes are often causal rather than logical consequences of care. The correlation between the care (structure and process) and the outcome may be very weak (32, 63). Another limitation is differences in the background population. This may be partly overcome by appropriate risk-adjustment (64-66).

Structural measures are rarely the focus of quality assessment and are often neglected completely despite the fact they form the framework for healthcare (50, 51). Along with structure, studying process measures should be paramount in high-quality care; it would be sensible to screen the process of care itself. However, there are several limitations to the use of process measures. First, only evidence-based care is suitable for evaluating whether "the right care" has been given, and only a small fraction of care is evidence-based (54, 67). Second, their use can be criticized for "what you measure is what you get" – the focus is placed on the measured processes while others are neglected (e.g. antibiotic prophylaxis at the cost of delayed onset of urgent operation) (68). Third, process measures may have a low public face validity since they involve intermediate steps instead of readily understood outcomes (62, 67). And last, like outcome indicators, also most process indicators require adjustment for case-mix (64).

1.4.3 Systems thinking

Systems thinking is a key element in quality improvement (7, 69). This is largely based on the Berwick's Central Law of Improvement: "Every system is perfectly designed to achieve the results it achieves." The core is that poorly designed systems, not individuals, are the ultimate causes of poor performance and poor outcomes (70) which rooted in Reason's theory of latent and active failure (70, 71). Reason suggested there are organizational structures and factors that have evolved over time and create "*error-producing and violation-promoting conditions*" (latent failure) even though the actual errors (active failure) are often committed by individuals.

Systems thinking was introduced in the IOM publication *To Err is Human* in 1999 (38). Thereafter systems thinking has overrun the prevailing idea of single cause framework of error in healthcare.

1.4.4 Near miss reporting – HaiPro in Finland

Blame is known to have a toxic effect on the quality of care by limiting open reporting and learning from poor outcomes and observed problems (38). In Finland, this issue was tackled by developing a specific web-based tool, HaiPro, for anonymous and voluntary reporting of adverse events and near misses. The sensitivity of these voluntarily reported incidents has been shown to be good and the data have been successfully used for quality improvements (72, 73).

HaiPro was developed at the Technical Research Center of Finland (VTT) in cooperation with healthcare units and was funded by the Ministry of Social Affairs and Health, National Agency for Medicines and VTT. It was piloted between 2007 and 2009 and today, HaiPro has over 144,00 users and is implemented in all hospital districts in Finland (72).

1.5 QUALITY OF CARE AND PATIENT SAFETY IN FINLAND

The Ministry for Social Affairs and Health issued the first national recommendation on quality management in social care and healthcare in 1995. The recommendation introduced three main principles for care providers: 1) quality management is part of everyday work, 2) client-orientation should be the core of quality management 3) social and healthcare quality management will be executed by knowledge management. The recommendation has later been updated with further emphasis on patient-centeredness, among other things(74).

In 2005, the Ministry for Social Affairs and Health launched an effort to coordinate and strategically guide patient safety activities at the national level. A steering group was established to examine and make development proposals on patient safety by the end of 2009. This work contributed to development of the patient safety reporting tool HaiPro. The national monitoring and reporting of quality of care and patient safety relies strongly on the National Institute of Health and Welfare (THL) and on the medical registers it maintains.

In 2010, the new Healthcare Act was issued and for the first time, the quality of care and patient safety were included among the main objectives of the act, the third of the five objectives being to "ensure universal access to the services required by the population and improve quality and patient safety" (75). This was further explained as "The provision of health care shall be based on evidence and recognized treatment and operational practices. The health care provided shall be of high quality, safe, and appropriately organized. ... Each health care unit shall produce a plan for quality management and for ensuring patient safety."

2 QUALITY OF CARE AND PATIENT SAFETY IN OBSTETRICS

2.1 QUALITY CHASM IN OBSTETRIC CARE

The "Bridges to Health" model by Lynn et al identified childbearing women and infants as one of eight population segments with distinct characteristics that must be addressed if the entire population is to achieve the IOM's aims for quality improvement (6). Globally, access to obstetric and perinatal care has improved but poor quality of the care is a major contributor to childbirth-related harm (76, 77). Despite the decreasing trend in maternal mortality (40% decrease globally between 1990 and 2015), preventable maternal deaths continue, especially but not solely in the systems with the scarcest resources. For example, in the USA half of the maternal deaths are estimated to be preventable (78-80). In addition, maternal mortality displays inequity and increasing racial disparities; In 2005, African-American women in USA experienced a rate of pregnancy-related death up to four times greater than for white women (81). In the UK, unequal access to antenatal care was associated with preventable maternal deaths (82).

In 2010, Ellsbury wrote about the quality chasm in neonatal-perinatal medicine providing a list of specific treatment practices that have been proven to be beneficial but are nevertheless not used (7). Ellsbury enlisted practices primarily aimed at improving the health of the neonate but the same problems concern many other obstetric practices; e.g. operative delivery, use of episiotomy and use of antibiotics for B streptococcal prevention may be overused for reasons not supported by clinical evidence (83-85).

There is variance in treatment practices between birth units and geographic areas, which are primarily due to differences in treatment culture and other extrinsic factors, not to differences in needs of mothers and newborns (7). The lack of a set of robust obstetric performance measures remains one of the main challenges in achieving standardized, evidence-based obstetric care (8).

In Finland, maternal and perinatal health is among the best in worldwide (25, 26). Maternal mortality ratio in 2015 was 3 per 100,000, whereas the average in the high-income countries was 10 (6 in Denmark, 4 in Sweden, 9 in the UK and 14 in the USA) (86). Reports and studies on regional differences are scarce, but e.g. the risk for obstetric trauma, a potentially preventable adverse event, has reported to vary significantly between birth units reflecting national differences in the quality of obstetric care (87).

Socioeconomic disparities have been reported to be low, but their impact on health status and access to healthcare services is increasing in Finland (88, 89).

2.2 HISTORY OF QUALITY IMPROVEMENT IN OBSTETRIC CARE

The concept of quality improvement emerged in obstetric and perinatal care in the 1970s. Reports of differences in neonatal mortality between obstetric facilities with and without neonatal intensive care units (NICU) in Canada inspired a collaboration towards integrated systems of perinatal care. This resulted in the first edition of *Toward Improving the Outcome of Pregnancy* (TIOP I) (16) by the March of Dimes in 1976 focusing on structural issues in perinatal and obstetric care. Around the same time, A. Cochrane awarded obstetrics with the notorious "wooden spoon" for making the worst use of evidence-based medicine (11). That was 30 years after the publication of the first medical RCT, and while by that time, many specialties had implemented the modern process of assessing the effectiveness of care, obstetric care still strongly relied on tradition and innovation rather than evidence (90). A decade later, in 1989, I. Chalmers and his group in Oxford came out with a collection of obstetric RCTs and published a massive two-volume work *Effective Care in Pregnancy and Childbirth* (91), marking the start of a new era of evidence-based obstetric care.

Already in the early 1990s, nearly a decade before the IOM publication *Crossing the Quality Chasm*, WHO and the European Union established an obstetric quality project (OBSQID) aiming at improving perinatal care through validated indicators (92, 93). This program was followed after 1999 by the ongoing Euro-Peristat project, using a pattern of indicators to monitor perinatal health in Europe (25). Since then, several other institutions with international impact, e.g. the Joint Commission, The Royal College of Obstetricians and Gynecologists (RCOG) and The Organisation for Economic Co-operation and Development (OECD) have provided their proposals for recommendable quality indicators within obstetrics (22, 94, 95).

In 2000, the United Nations included improving maternal health and reducing child mortality among the eight Millennium Development Goals. This had a substantial impact on the international interest on the quality of obstetric and perinatal care, resulting in significant developments in maternal and neonatal health in the low- and middle-income countries (96-98). In 2012, WHO presented the Safe Childbirth Checklist, with promising results on improving perinatal outcomes (76). These actions by the worldwide, high-impact organizations signaled emerging concerns about the equitability, safety, and effectiveness of perinatal care, but also, they direct

public attention to the matter and provide concrete tools for quality improvement.

Finland has an extensive history of detailed civil registries and there are recordings of perinatal health data (mortality) already from 1800 onwards (99). Today, the national MBR, maintained by THL, systematically gathers, revises and publishes perinatal health data covering virtually 100% of births in Finland (29). However, despite the vast amount of high-quality perinatal health data gathered nationally, a culture of local, systematic and continuous quality monitoring is still lacking. Recently, THL published a recommendation on a set of indicators that would be beneficial for continuous monitoring and quality improvement (100). In Denmark and Sweden, such national quality monitoring systems providing regional and/or local data on different outcomes (structure/process/outcome indicators as well as costs) have been used for several years; the Danish National Indicator Project (DNIP) and the "Open Comparisons" (Öppna Jämförelser) in Sweden enable benchmarking between units and facilitate quality improvement by learning from the differences (101, 102).

2.3 UNIQUE CHARACTERISTICS OF OBSTETRIC AND PERINATAL CARE

2.3.1 Maternal and neonatal needs are not always aligned

Although maternal and neonatal health is no doubt coupled, the best care for the mother may compromise the optimal care of the neonate, and vice versa. An intervention, such as labor induction in prolonged pregnancy, could potentially improve neonatal outcomes, but simultaneously increase maternal risks by increasing the risk for operative delivery. This sets special requirements for the interpretation of the research evidence, and furthermore, for creating the guidelines for best practice.

High-quality obstetric care needs to be driven by balancing the best interest of both, the mother and the neonate, and this requires collaboration between neonatologists and obstetricians. Historically, however, that generally has not been the case but the increasing focus on quality improvement in recent years has fostered a more cohesive team approach (103).

2.3.2 Medicalization of a physiological process

Obstetric care is not focused on treating pathology but, essentially, on preventing harm during one of the most natural but yet most hazardous events of life. This sets care providers with a unique challenge – how to

ensure maximal safety while simultaneously providing only the effective care, refraining from any unnecessary medical interventions.

Already in 1985, WHO claimed that the entire modern obstetric and neonatologic literature was based on observations of 'medicalized' birth and that most healthcare providers no longer understood the idea of a 'non-medicalized' birth (104). In balancing efficacy and risks, care providers tend to resort to efficacy. This is often also the focus of research setup and study hypothesis, leading to rapid implementation of new, proactive treatment protocols – often unnecessary or even harmful at the population level (21). Globally, medicalization of birth is a more significant concern than in Finland (21), and it is counterbalanced by the public's increasing awareness and interest in natural childbirth (105).

Fear of childbirth has been associated with birth medicalization (106). The theory is that, with the increasing intervention rates, the public has been unintentionally educated to fear natural birth and its rare complications and simultaneously to foster overoptimistic notions of care providers' capabilities to manage whatever complications the parturient woman encounters. Fear of childbirth is a well-recognized obstetric problem also in Finland, leading to potentially avoidable Cesarean sections (107).

2.3.3 Long-term impact on population health

From a subjective perspective, the care given around birth has a significant impact on the well-being of a family and a deep, long-term impact on women's perceptions about their bodies and their babies' capabilities (108). At a population level, pregnancy and birth offer a poignant possibility to assess and impact long-term the health of a society through numerous different social and physical determinants of health – not only by ensuring healthy newborns and future citizens but e.g. by optimal care of gestational diabetes, and education on its impact later in life (109).

2.4 ASSESSMENT OF QUALITY AND PATIENT SAFETY IN OBSTETRIC CARE

2.4.1 Choosing obstetric quality indicators

There is an abundance of different obstetric quality indicators tested and used (51); the enduring dilemma is how to choose and implement a set of standardized indicators that would be comparable over time and between different units (agreed-on risk-adjustment methods), and preferably suitable also for international comparisons (110, 111). National validation is known to be essential for quality indicators, but nationally available data are flawed and limited (62, 112). In Finland, we lack both a national consensus on the obstetric quality indicators and a national validation of the existing indicators.

As discussed above in the first section of literature review, the Donabedian model of measuring healthcare quality in three dimensions has directed the development of quality indicators in nearly all medical specialties, also within obstetrics (46). Enlisted are examples of quality indicators categorized according to the Donabedian model:

- The ratio between midwives and women in a delivery unit (structure)
- Time from decision to delivery by emergency CS (process)
- Obstetric trauma (outcome)

IOM, in turn, introduced six dimensions, or domains, of quality (safety, effectiveness, patient-centeredness, timeliness, efficiency and equitability). These can be regarded as counter-balancing aims for high quality healthcare, e.g. maximally safe care is not necessarily maximally effective care (4). An ideal set of quality indicators would measure most, if not all, IOM domains of quality and target all (Donabedian) dimensions of care.

Most indicators can be distinguished as rate-based or sentinel (50). The rate-based indicators, e.g. emergency CS rate or admission to NICU, can be used in quantitative analysis of quality. They are mostly adverse events or practices that should be avoided but possess a rate below which the quality is not necessarily improved. Finding the “optimal” incidence is a complex task but continuous monitoring and benchmarking will likely contribute to determining the cut-off rate (113).

The sentinel indicators, in turn, are serious but infrequent adverse events that should always trigger further analysis, the most extreme examples being maternal and neonatal deaths. Although these are important to track, they may perform poorly as quality indicators due to their insensitivity (24). Classifying between sentinel and rate-based indicators is not clearly categorical, e.g. peripartum hysterectomy can be regarded as a sentinel event as in outcome of a poorly managed postpartum bleeding, but it is also a life-saving procedure in a serious condition like placenta perccreta. In the latter

case, the indicator could be regarded rate-based – it does not reflect inadequate care but different essential (structural) quality aspects of care such as appropriate access to the operating room and sufficient skills of the attending physician.

2.4.2 Obstetric and perinatal quality indicator series in clinical use

Tables 2 and 3 present the quality indicators used by several internationally acknowledged institutions and organizations as well as the Nordic quality systems of DNIP (Denmark) and Open comparisons (Sweden), and a proposal by THL for quality indicators. An important feature of a reliable quality indicator is local/national validation (50, 114) - the Nordic projects are presented as examples of locally validated indicators from systems largely similar to ours.

The indicators are categorized by applying the Donabedian dimensions. In addition, the tables include one interpretation of which aspects of the care each indicator measures according to the IOM dimensions/domains. Compiled, this classification by outcome, process, and structure indicators highlights the safety orientation of current obstetric quality measurement. The individual indicators presented in the tables are discussed more in detail in the following sections. Composite indicators for quality measurement are covered in the final section. Mostly, they are not implemented in national quality programs, and therefore, are not included in the tables.

2.4.3 Outcome indicators

Most outcome indicators attempt to measure the quality through adverse outcomes; the focus is on safety.

Obstetric trauma is one of the few internationally recognized and accepted obstetric quality indicators, recommended by e.g. the OECD, AHRQ, and RCOG (1, 22, 95). Its pros include the high face validity in that it seems a relevant indicator due to clear evidence of the trauma's negative impact, and to it being potentially preventable (115). Its use as an indicator is enhanced by relatively objective diagnostics and coding (116) and low cost and low burden tracking through hospital administrative systems (117).

Table 2 *Outcome indicators for quality of obstetric and perinatal care.*

OUTCOME INDICATOR	SPECIFICATIONS	DIMENSION	INSTITUTION
MATERNAL MORTALITY	Overall	Safety	THL
	By cause of death	Safety	Euro-Peristat
	Following a complication	Safety	
OBSTETRIC TRAUMA	In all vaginal deliveries	Safety	Euro-Peristat, RCOG
	On nulliparous	Safety	Open Comparisons
	In non-instrumental and instrumental vaginal deliveries		RCOG, OECD, AHRQ
SEVERE MATERNAL MORBIDITY	Uterine rupture	Safety	THL
	Peripartum hysterectomy	Safety	THL
	Postpartum bleeding: - ≥1,000 mL - with coagulation defects		THL
	Blood transfusion (process/outcome)	Safety, Equity	THL
NEONATAL MORTALITY	Perinatal mortality	Safety	THL
	Neonatal mortality	Safety	Open Comparisons
	Stillbirths	Safety	Euro-Peristat, Open Comparisons
NEONATAL HYPOXIA	Low uc-pH and/or Apgar score	Safety	THL, DNIP
	pH <7.05	Safety	THL
	5-min Apgar <7 (and <4)	Safety	THL, Open Comparisons, Euro-Peristat
NEONATAL INFECTION	Newborn bloodstream infections	Safety	The Joint Commission
	Infections in neonatal ward	Safety	Open Comparisons
BIRTH TRAUMA ON NEONATE		Safety	AHRQ
ADMISSION TO NICU (STRUCTURE/PROCESS/OUTCOME)	>2,500 g and for > 24 hours†	Safety	
VISUAL ANALOG SCALE (VAS)	Birth experience by the mother	Efficiency, Safety, Patient-Centeredness, Equity	Finland (in clinical use)

Table 3 *Process and structure indicators for quality of obstetric and perinatal care*

PROCESS AND/OR STRUCTURE INDICATOR	SPECIFICATIONS	DIMENSION	INSTITUTION
MODE OF DELIVERY (OUTCOME/PROCESS)	by parity, plurality, presentation (of fetus), previous CS	Effectiveness	Euro-Peristat
	Unassisted vaginal delivery	Effectiveness, Patient-centeredness, Efficiency	RCOG, Euro-Peristat
	Instrumental delivery rate	Safety, Effectiveness	RCOG
	Vaginal birth after CS (VBAC)	Safety, Effectiveness	AHRQ (IQI 34)
CESAREAN SECTION (CS) (OUTCOME/PROCESS)	Overall	Safety, Effectiveness	RCOG, The Joint Commission (Euro-Peristat)
	Pre-labor CS		RCOG
	CS rate in low-risk groups: - Primary, uncomplicated - On nulliparous (with term, single cephalic pregnancy)	Safety, Effectiveness	AHRQ (IQI 33), Open Comparisons, Joint Commission
	Emergency CS rate: - in spontaneous labor - following labor induction	Safety, Effectiveness	RCOG
	CS performed under general anesthesia	Effectiveness, Safety	DNIP, THL
	Emergency CS within time recommended from decision to birth	Safety, Timeliness	DNIP
ONSET OF LABOR (PROCESS)	Distribution of births by mode of onset of labor	Effectiveness	Euro-Peristat
	Pre-labor delivery	Effectiveness	Joint Commission
	Proportion of labor induction	Safety, Effectiveness	RCOG
	Labor induction between 37 and 39 weeks and pre-labor CS before 39 weeks (no clinical indication)	Safety, Effectiveness	RCOG
	Labor induction after ≥42 weeks		RCOG
EPISIOTOMY RATE (PROCESS)	Overall	Effectiveness	Euro-Peristat
	In instrumental and non-instrumental vaginal deliveries	Effectiveness	RCOG
ANESTHESIA / PAIN RELIEF (PROCESS/STRUCTURE)	Proportion of birth, epidural/birth spinal given within one hour from prescription.	Patient-centeredness, Effectiveness	DNIP
READMISSION	Emergency maternal readmission within 30 days of delivery	Safety, Effectiveness, Efficiency	RCOG, THL
CORD BLOOD PH (UC-PH) MEASUREMENT (PROCESS/STRUCTURE)	Proportion of newborns with uc-pH taken	Efficiency, Equity	THL
CARE OF PRETERM DELIVERY (PROCESS/STRUCTURE)	Antenatal steroids in pre-term delivery	Safety, Effectiveness	Joint commission 2015
	Percentage of very preterm infants delivered in units without a NICU	Safety, Equity	Euro-Peristat
ESTABLISHMENT OF SKIN-TO-SKIN CONTACT (PROCESS)	Between mother and newborn infant	Effectiveness, Patient-centeredness	DNIP, Joint Commission
BREASTFEEDING (PROCESS)	Percentage of infants breastfed at birth	Effectiveness, Patient-centeredness	Euro-Peristat
	Exclusive breastfeeding		Joint commission
	Exclusive breastfeeding, considering mother's choice		Joint commission
"OPTIMAL DELIVERY" (OUTcOME/PROCESS)	Delivery of a healthy child after uncomplicated delivery	Effectiveness, Safety	DNIP
	Births without obstetric intervention		Euro-Peristat
CONTINUOUS SUPPORT (PROCESS, STRUCTURE)	for women in delivery room	Efficiency, Safety, Patient-Centeredness, Equity	DNIP
PLACE OF BIRTH (STRUCTURE)	Distribution by volume of deliveries	Safety, Effectiveness	Euro-Peristar
	Unplanned out-of-hospital delivery	Safety	THL
COST OF BIRTH		Efficiency	Open Comparisons

Obstetric trauma refers to obstetric anal sphincter injuries encompassing both 3rd and 4th degree perineal lacerations. The 3rd degree injury is defined as partial or complete rupture of the anal sphincter muscles (the internal or the external or both), the 4th degree involves also the rectal mucosa. It is a serious, potentially preventable complication of vaginal delivery and may result in significant and long-term symptoms like perineal pain and anal incontinence (118-120).

There are very impressive results on clinical programs aiming to decrease the incidence of obstetric trauma corroborating the previous evidence of the trauma being preventable by appropriate labor management (121-123). Also, there are indications that merely the tracking of the incidence would decrease it (124). This supports the concept of continuous monitoring as part of quality improvement – awareness and reporting may independently contribute to improvement without additional efforts. Obstetric trauma is, however, a simple outcome indicator and as such, significantly influenced by patient characteristics. Therefore, the indicator requires risk-adjustment to stand inter-hospital comparisons, and somewhat arbitrarily, its use as a quality indicator has been criticized for this reason (125). The variation in background population is, after all, the main problem to overcome in quality analysis.

The neonatal outcome indicators used in quality monitoring are largely based on fetal hypoxia indicators; Apgar score (126, 127) and umbilical cord pH (uc-pH) (128). Both have their limitations, Apgar score is based on subjective evaluation, whereas uc-pH is objective, but the routines in how often uc-pH is tested and reported are variable (100).

Birth trauma, an AHRQ safety indicator, comprehends several independent traumatic conditions of the neonate (1). The indicator's usability in obstetric quality improvement has been questioned due to lack of construct validity; the indicator does not seem to be associated with other accepted quality measurements, e.g. prolonged stay of the neonate (129). The indicator is defined according to the ICD-9-CM codes (130), which are designed for utilization in the United States only and not used in Europe. This further complicates the indicator's use in international settings, and thus far, it has not been validated or implemented in clinical use in Europe.

The Vermont Oxford network provides an example of international surveillance system based on neonatal outcomes; the network aims at evidence-based quality improvement in neonatology (131). Established in 1988 as a collaboration of 34 NICUs gathering and sharing data on outcomes on very low birth weight infants, the network has grown into a database of over 1,000 units holding information on more than 2.2 million newborns. The data are used for benchmarking and improving treatment protocols internationally (132, 133). The focus is on high-risk group of preterm neonates, but similar systems should perform well for low-risk groups, too.

2.4.4 Process and structure indicators

2.4.4.1 Cesarean section rates and the use of the Robson classification

Rising CS rates continue to be a major concern globally, and attempts to define 'optimal rate' at a population level have failed (134-137). Already in 1985, WHO stated that CS rate should not exceed 15% in any population. Yet, thirty years later, the world average CS rate was 19% and for developed countries 27% (26, 138). In 2015, after the dramatic increase in CS rates over the past decades, WHO issued a statement where it refrained from presenting any target CS rates, but emphasized that the focus should be on providing a Cesarean section to every woman in need (139). In the developed world, where CS is widely available, the problem is how limit its use solely to the women who will benefit from the intervention. Unnecessary CS has been presented as the quintessential example of medicalization and dehumanization of birth (21).

Total CS rate, and different CS rate subclasses like pre-labor CS and CS on low-risk women, are known obstetric quality indicators (1, 23, 95, 101, 102, 140). Validity is a concern, especially when using the overall rates (141).

The Robson classification system, also known as the ten-group classification system (TGCS) is a tool for stratified CS rate analysis (142). It is a totally inclusive and mutually exclusive woman-based classification system founded on the idea that outcome rates (primarily CS rates) within each of the classification groups are comparable over time and between different units/populations. The system is recommended by WHO to be used as a global standard for assessing, monitoring, and comparing CS rates within and between healthcare facilities (139). It is considered useful both for monitoring CS rates and for reducing them when necessary (143-148).

The classification is based on a few simple obstetric parameters: number of fetuses, fetal presentation, gestational age (GA), previous CS, parity, and onset of labor, resulting in 10 separate groups (Table 4). There is a strict hierarchy in the classification process illustrated in Figure 1. The stepwise classification ensures that all parturient women fall into one and only one category (totally inclusive and mutually exclusive). Discarding the hierarchy will cause misclassification – a known hazard for Robson group studies (149). For example, the number of fetuses and fetal presentation are superior to GA threshold, and therefore, the groups R6-R9 contain all multiple pregnancies and all pregnancies with non-cephalic presentations, also preterm and ones with a previous CS. By contrast, the group 5 (R5) contains solely the women with a previous CS and a single, term, cephalic pregnancy (not breech, not preterm).

Table 3 *The 10 Robson groups according to the classification system (142)*

R1	Nulliparous, single cephalic, ≥ 37 weeks, in spontaneous labor
R2	Nulliparous, single cephalic, ≥ 37 weeks, induced (2a) or pre-labor CS (2b)
R3	Multiparous (excluding previous CS), single cephalic, ≥ 37 weeks, in spontaneous labor
R4	Multiparous (excluding previous CS), single cephalic, ,≥ 37 weeks, induced (4a) or pre-labor CS (4b)
R5	Previous CS, single cephalic, ≥ 37 weeks
R6	All nulliparous with breech presentation
R7	All multiparous with breech presentation (including previous CS)
R8	All multiple pregnancies (including previous CS)
R9	All abnormal lies (including previous CS)
R10	All single cephalic, ≤36 weeks (including previous CS)

In a systematic review assessing different classification systems for CS, the woman-based systems proved superior to urgency- or indication-based systems, with the Robson classification attaining the highest score (148). The method of classifying "who" (woman-based), instead of "why" (degree of urgency) or "when" (indication) was considered conceptually easy, to have good reproducibility, to be well tested, and to perform well in prospective settings. The lack of information on the indication was the only flaw of the Robson classification system brought up in the study.

While the Robson classification is widely regarded as the golden standard for stratified CS rate analysis today, it has some limitations. In a systematic review of the Robson classification, the limitations that emerged were as follows: 1) lack of information on the indication 2) the residual heterogeneity within the groups (including pre-existing medical conditions, maternal age and BMI), 3) the need to adjust for heterogeneity in inter-hospital comparisons, 4) the classification not being able to evaluate the direct relationship between CS and outcomes, and 5) resources needed for implementation (149)

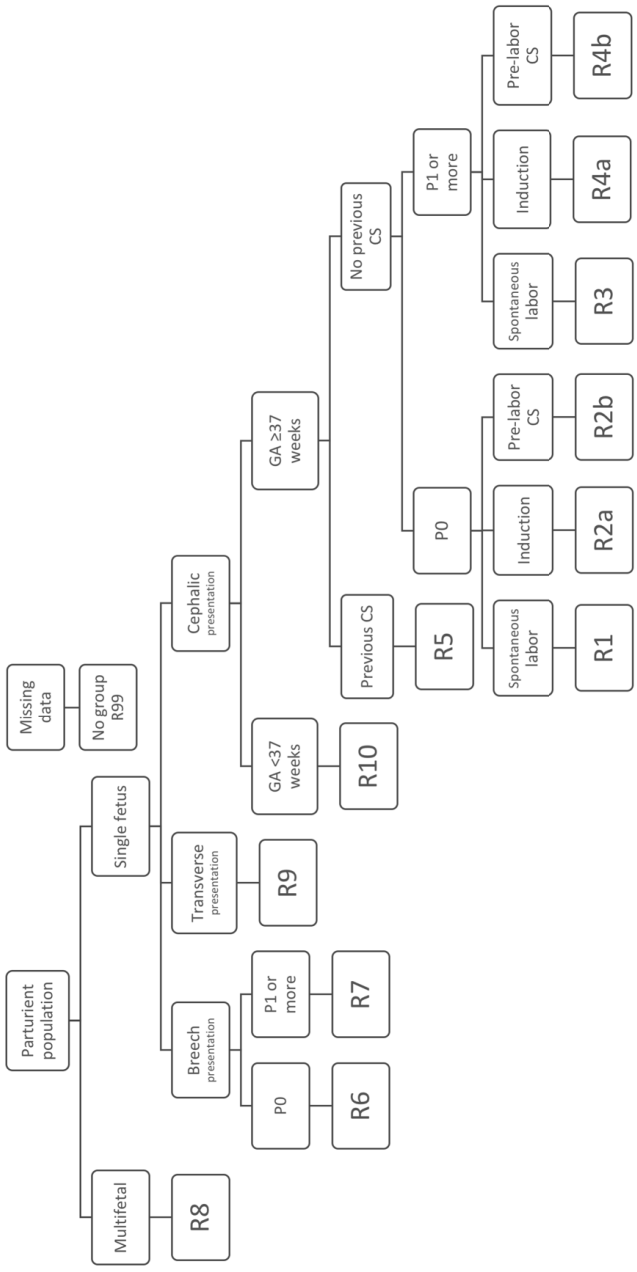


Figure 1 *Flow chart of the Robson classification process*

2.4.4.2 Size of a birth unit

Previously, it has been shown that concentrating high risk deliveries (GA less than 32 weeks or expected birth weight <2,000 / 1,500 grams) to tertiary care units with NICU decreases neonatal mortality (15, 150, 151). It is crucial that the obstetric volume of these units is high enough, since too few births per NICU lead to less effective care (150). The large units have also been associated with fewer patient safety claims considered to reflect the quality of care (152). On the other hand, birth regionalization to fewer but larger units has been criticized for increasing the risk for unplanned out-of-hospital deliveries, and therefore increasing perinatal risks (19).

How the regionalization affects the quality of low-risk maternal care has been studied little (20). Medicalization of birth seems immediately connected to the expanding obstetric volumes and is accompanied by increased operative delivery rates and decreased rates of births without intervention (21).

2.4.4.3 Novel process and structure measures

Many of the process measures presented in Table 3 represent the traditional idea of a process indicator, e.g. induction and operative delivery rates. However, these are often outcomes or end-results of underlying processes or patient characteristics, and they are similarly confounded as the outcome indicators. Implementing broader quality measures that would contribute to improving systems of care has remained a challenge (62).

There are some novel indicators like establishment of skin-to-skin contact between mother and newborn infant, which are readily applicable and understandable and most importantly, supported by strong evidence (24). Continuous support for women in the delivery room is shown to reduce the intervention rates and is therefore closely related to patient safety (24). However, the indicator could be considered prone to problems with definitions of “continuous support”: How strictly should this be defined? From how early in the labor is the support expected? These two indicators, as well as, “delivery of a health child after uncomplicated delivery”, reflect several aspects of the structure and process of care and do not readily fit the traditional Donabedian model. They seem plausible quality indicators, but due to their comprehensive nature, require local validation and careful interpretation.

2.4.5 Composite indicators: Adverse outcome indices, Bologna score, and visual analog scale

In the USA, the lack of a nationally accepted obstetric indicator series was tackled by organizing two large consensus conferences through which an Adverse Outcome Index (AOI) was developed (111). AOI is based on the following ten adverse events that are tracked (in parenthesis the weighing score for each event to balance the adverse events based on the severity for a weighted adverse outcome score, WAOS):

- Maternal death (750)
- Intrapartum or neonatal death (400)
- Uterine rupture (100)
- Maternal admission to ICU (65)
- Birth trauma (on neonate) (60)
- Return to operation room (40)
- Admission to NICU (over 2,500g and for over 24h) (35)
- 5-min Apgar <7 (25)
- Maternal blood transfusion (20)
- Obstetric trauma (5)

AOI and WAOS offer one perspective from which adverse outcomes should be monitored, and to prioritized. However, compared with the indicators proposed and used nationally or by larger institutions (Tables 2 and 3), the components of AOI seem somewhat simplistic, partly due to the more limited data sources at the time of its development. In addition, the tool completely ignores some essential elements like mode of delivery, as well as many structural elements that, in recent years, have been brought into the core of obstetric quality assessment (153, 154). Similar to AOI, a composite indicator focusing on neonatal adverse events only was developed in Australia (neonatal adverse outcome indicator, NAOI) (155).

The usability and efficiency of AOI/WAOS and NAOI in obstetric quality improvement have been both appraised and questioned but overall, they have not been widely implemented in clinical use (155-157). The main problem with adverse event indices is related to weighting schemes. A weighting system subjects the indicator to a more complicated interpretation which, in turn, decreases the face validity and does not contribute to simple and straightforward monitoring systems (158).

Bologna score offers a different approach from the previous systems; instead of adverse events, the Bologna score aims at assessing effective management of uncomplicated deliveries (159). It is based on five, non-weighted yes/no measures: 1) presence of a companion at birth 2) use of a partogram 3) absence of augmentation (use of oxytocin) 4) use of a nonsupine position at birth 5) skin-to-skin contact between the mother and neonate (for at least 30 minutes within the first hour after birth). The maximal score of 5 is considered to represent effective management of

normal labor, and scores closer to 0 indicate either less optimal management of normal labor or variation from normal labor itself. Therefore, the indicator is very prone to variability in the background factors of the parturient population (measuring how normal the labor was, not how optimally it was managed). Similar to the Bologna score, the DNIP indicator "Delivery of a health child after uncomplicated delivery" and the Euro-Peristat indicator "Births without obstetric intervention" aim to measure the optimality of labor management (24).

A patient experience measurement, the visual analog scale (VAS) is acknowledged here as a composite indicator due to the comprehensive nature of VAS as a quality indicator. Since 2005, VAS has been used in most birth units in Finland and Sweden for measuring the parturient woman's birth experience (in addition to labor pain). It is incorporated in the electronic medical record system, Obstetrix, and therefore easy to track (160). The indicator is especially valid for screening potentially traumatizing birth experience and women in risk for developing childbirth-related fear (161).

2.5 BEST PRACTICE PROTOCOLS FOR IMPROVING QUALITY OF OBSTETRIC CARE

2.5.1 National guidelines and best practice protocols

National guidelines and best practice protocols have significant impact on clinical practices. Common guidelines are crucial for equitable care but the ultimate problem is defining the best practice, the golden standard, when there is insufficient or conflicting evidence (24). Vaginal birth after Cesarean section (VBAC) rates in USA offer an example of fluctuating evidence, recommendations and practices: Guideline updates from the American College of Obstetricians and Gynecologists (ACOG) resulted in VBAC rates from around 3% in 1970s to nearly 30% in 1995 and back to less than 10% in 2005(162). A different example is breech deliveries: the conclusions of a single research paper profoundly changed clinical practice resulting in dramatic decreases in vaginal breech deliveries (163-165).

2.5.2 Timing of labor induction

The effect of labor induction compared with expectant management has been extensively studied, but the results are partly conflicting, and a consensus on the optimal time of induction in late-term (41^{+0} and 41^{+6}) and postterm (42^{+0} gestational weeks and beyond) pregnancy is lacking. Systematic reviews on the topic are mostly in favor of active management of prolonged pregnancies, with respect to both the risk for CS (28, 166, 167) and the risk for adverse neonatal outcomes (28, 168), or indicate a similar risk in labor induction versus expectant management (27). In most studies, induction is recommended after 41 weeks.

Cohort studies, especially those comparing labor induction with spontaneous labor, show higher risk for adverse outcomes in induced deliveries, especially for nulliparous women (169-173). However, instead of spontaneous labor, labor induction should be compared with expectant management (174). In such cohort setups the adverse outcome rates seem to be similar in both cohorts (175, 176), or even superior in the later induction cohort (177).

2.5.3 Risks associated with postterm pregnancy

The prevalence of postterm pregnancy in a given population is dependent on population characteristics and obstetric management, including the accuracy of pregnancy dating. Due to active management of prolonged pregnancy, estimating how many of the induced pregnancies would continue beyond 42 gestational weeks is nearly impossible. In Finland, a widely accepted protocol

has been to induce labor at 42^{+1} unless prior clinical examination (usually performed at 41^{+5} by an obstetrician) arouses concerns about the well-being of the fetus (178). This is a rather conservative approach, which is changing towards more proactive policies (induction before 42 weeks). The prevalence of postterm pregnancies in Finland (4.7% in 2010) is likely close to the true incidence (25). In USA, the prevalence was 5.6% and in Europe, the national prevalence was between 0.1% (Malta) and 6.6% (Sweden) in 2010 (25, 179).

Risk factors for postterm pregnancy are not clearly established, but it is known that previous postterm pregnancy, male fetus, nulliparity and obesity increase the incidence (180-184). The effect of obesity has been explained by genetic factors and hormonally active adipose tissue altering the hormonal status (185, 186). In Finland, the proportion of nulliparous deliveries has remained steady (28-29%) and the maternal BMI has increased only slightly (proportion of parturients with BMI over 25 from 32% to 35% between 2008 and 2015).

Considering that in Finland more than one out of five births occur after 41^{+0}, a change towards an induction policy close to 41 weeks would have a significant impact on the induction rates which are increasing as it is (29). Following an international trend, our national induction rate has increased from 16.6% to 24.8% between 2005 and 2015 and interestingly, this has resulted in only marginal changes in the prevalence of late-term (18% vs 19%) and postterm (4.9% vs 4.0%) births (29). The increase in the induction rates may be partly explained by the improved recordings of labor inductions. It is noteworthy that the total CS rate has slightly decreased between 2005 and 2015 (16.5 - 15.9%) in Finland. However, this is mainly due to decreased pre-labor CS rates (7.3 - 6.1%) while during the increase in induction rates, the emergency CS rate increased slightly (9.2 – 9.7%) (29).

In Denmark, after the implementation of a new practice protocol for prolonged pregnancies, the postterm birth rate decreased from 5.6% to 1.5% (late-term birth rate increased from 17.5% to 22.5%) and this change has been reported to be associated with improved perinatal outcomes (187). Meanwhile, the percentage of births with spontaneous onset decreased from 78% to 68% between 2006 and 2012 (respective change in Finland from 76% to 73%) (29, 187). The rate of pre-labor CS and emergency CS remained steady, but was significantly higher than in Finland (9.4% – 9.8% and 10% – 10%, respectively)(187).

Due to the versatile and vast obstetric population with prolonged pregnancies, RCTs assessing the problem need to cover a large population, and the study setup must be carefully planned. A registry-based setup, covering the whole parturient population and equipped with adequate risk-adjustments and case-control comparisons, would potentially enable determination of the optimal time of delivery.

AIMS OF THE STUDY

Aims of the individual studies of this thesis were as follows:

1) To analyze the trends in obstetric care by a stratified Cesarean section analysis in the Nordic countries using the Robson classification (I)

2) To analyze differences in adverse obstetric and neonatal outcomes of obstetric care by the size of the delivery unit in Finland (II-III)

3) To evaluate and validate different obstetric and neonatal outcome and process measures for quality assessment (II-III)

4) To determine the optimal time of delivery in prolonged pregnancy (IV)

MATERIALS AND METHODS

3 STUDY POPULATIONS

3.1 CESAREAN SECTION TRENDS IN THE NORDIC COUNTRIES (I)

This international study included all births (n=3,398,586) in the Nordic countries between 2000 and 2011: 757,257 (22.3%) from Denmark, 690,144 from Finland (20.3%), 52,607 (1.5%) from Iceland, 699,754 (20.6%) from Norway and 1,198,824 (35.3%) from Sweden. Explicitly births, not women, were included and hence all women were included by the number of times they gave birth during the study period. There were no exclusion criteria for this study.

Characteristics of the study population are shown by country in Table 5.

3.2 OBSTETRIC TRAUMA AND NEONATAL OUTCOME AND PROCESS MEASURES (II, III)

The original study population for these studies consisted of all births (n=294,725) in all Finnish birth units (n=34) between the years 2006 and 2010. How the individual study populations were constructed is shown in the flow diagram in Figure 2.

For the study on obstetric trauma, the outcomes were assessed separately for all births (n=294,725) and subpopulations of all vaginal births, instrumental vaginal births and non-instrumental vaginal births as recommended by OECD (22)..

In the study on neonatal outcome, the aim was to focus on low-risk births. Therefore, the study population was limited to term, singleton deliveries (beyond 37 gestational weeks). By excluding university clinics, we excluded most of the high-risk pregnancies and eliminated the confounding factor of the known superior critical care resources in the university clinics. We performed separate analyses for deliveries with a GA of 42 weeks or more (n=15,020) because they are known to be associated with higher neonatal morbidity and complication rates (188, 189).

Table 4 *Characteristics of the Nordic Study population (I)*

		Denmark	Finland	Iceland	Norway	Sweden	Total
Births n (% of total)		757,257 (22.3)	690,144 (20.3)	52,607 (1.5)	699,754 (20.6)	1,198,824 (35.3)	3,398,586 (100)
Cesarean section n (CS rate %)		145,655 (19.2)	113,202 (16.4)	8,747 (16.6)	109,690 (15.7)	200,989 (16.7)	578,283 (17.0)
Parity n (% of births	nulliparous	318,656 (42.1)	288,635 (41.8)	20,988 (39.9)	290,589 (41.5)	551,294 (46.0)	1,470,162 (43.3)
	multiparous, no previous CS	352,423 (46.5)	328,296 (47.6)	25,812 (49.1)	347,455 (49.7)	545,062 (45.5)	1,599,048 (47.1)
	multiparous, previous CS	86,178 (11.4)	72,511 (10.5)	5,807 (11.0)	61,710 (8.8)	102,468 (8.5)	328,674 (9.7)
	missing data	0 (0.0)	702 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	702 (0.0)
Presentation n (% of births	cephalic	721,195 (95.2)	665,494 (96.4)	50,745 (96.5)	650,265 (92.9)	1,154,619 (96.3)	3,242,318 (95.4)
	breech	33,057 (4.4)	20,907 (3.0)	1,662 (3.2)	24,747 (3.5)	43,175 (3.6)	123,548 (3.6)
	oblique or transverse	2,186 (0.3)	3,743 (0.5)	159 (0.3)	2,054 (0.3)	1,030 (0.1)	9,172 (0.3)
	missing data	819 (0.1)	0 (0.0)	41 (0.1)	22,688 (3.2)	0 (0.0)	23,548 (0.7)
Multiple pregnancy n (% of births)		21,558 (2.8)	10,454 (1.5)	933 (1.8)	12,450 (1.8)	17,457 (1.5)	62,852 (1.8)
Gestational age n (% of births)	<37 weeks	45,358 (6.0)	35,883 (5.2)	2,689 (5.1)	43,255 (6.2)	66,722 (5.6)	193,907 (5.7)
	≥ 37weeks	707,848 (93.5)	651,948 (94.5)	49,887 (94.8)	650,993 (93.0)	1,132,102 (94.4)	3,192,778 (93.9)
	missing data	4,051 (0.5)	2,313 (0.3)	31 (0.1)	5,506 (0.8)	0 (0.0)	11,901 (0.4)
Onset of the labor n (% of births)	spontaneous	575,041 (75.9)	526,464 (76.3)	39,888 (75.8)	541,798 (77.4)	95,8671 (80.0)	2,641,862 (77.7)
	induced	104,146 (13.8)	115,457 (16.7)	8,734 (16.6)	9,9975 (14.3)	140,304 (11.7)	468,616 (13.8)
	pre-labor CS	78,070 (10.3)	48,223 (7.0)	3,546 (6.7)	57,981 (8.3)	99,849 (8.3)	287,669 (8.5)
	missing data	0 (0.0)	0 (0.0)	439 (0.8)	0 (0.0)	0 (0.0)	439 (0.0)
Mode of delivery n (% of births)	vaginal cephalic	549,830 (72.6)	524,184 (76.0)	40,017 (76.1)	529,807 (75.7)	909,139 (75.8)	2,552,977 (75.1)
	vaginal breech	288 (0.0)	505 (0.1)	418 (0.8)	8,983 (1.3)	3,000 (0.3)	13,194 (0.4)
	instrumental vaginal	61,538 (8.1)	51,458 (7.5)	3,417 (6.5)	51,274 (7.3)	85,696 (7.1)	253,383 (7.5)
	emergency CS	81,994 (10.8)	48,223 (7.0)	3,352 (6.4)	41,828 (6.0)	99,849 (8.3)	275,246 (8.1)
	pre-labor CS	63,607 (8.4)	64,979 (9.4)	5,395 (10.3)	67,862 (9.7)	101,140 (8.4)	302,983 (8.9)
	missing data	0 (0.0)	795 (0.1)	8 (0.0)	0 (0.0)	0 (0.0)	803 (0.0)

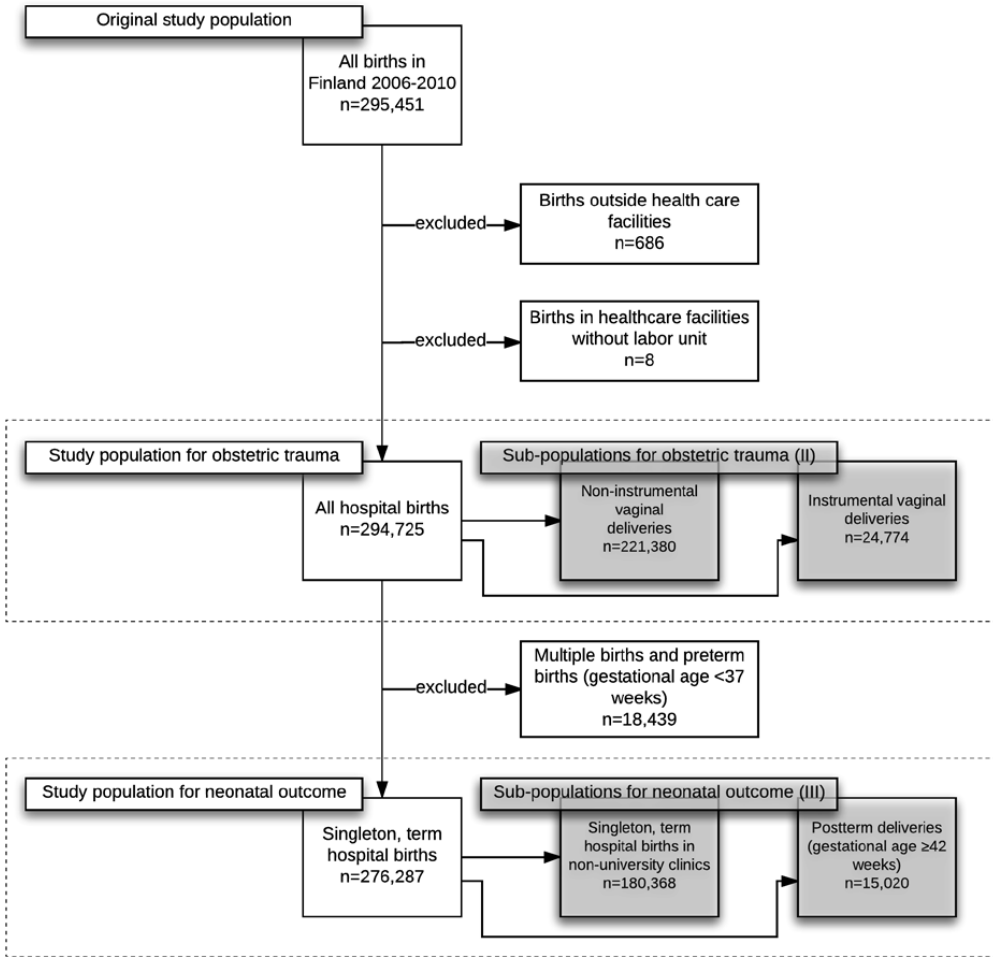


Figure 2 *Obstetric trauma and neonatal outcomes in the Finnish birth units: Flow chart of the study population (II, III)*

3.3 EFFECT OF LABOR INDUCTION IN PROLONGED PREGNANCY (IV)

The original population consisted of all deliveries in Finland between 2006 and 2012 (n=420,061). There were 212,716 singleton cephalic deliveries with GA between 40^{+0} and 42^{+3} included in the study population (nulliparous 44%, n=92,812). Exclusion criteria are shown in the flow chart in Figure 3. The proportion of postterm (GA 42^{+0} or more) deliveries was 4.8% (n=20,194) during the study period

The pre-labor CS cases were excluded from the study population (n=2,998, 1.4% of total), and this was acknowledged as a potential confounder. Excluding them was a deliberate choice based on the presumption that the majority of the pre-labor CS cases would not have been eligible for labor induction at any GA beyond 40^{+0}. When performed after 40^{+0}, there is, in most cases, a clear pre-existing indication for elective CS (e.g. macrosomia or strong fear of birth or labor induction) that contraindicates labor induction.

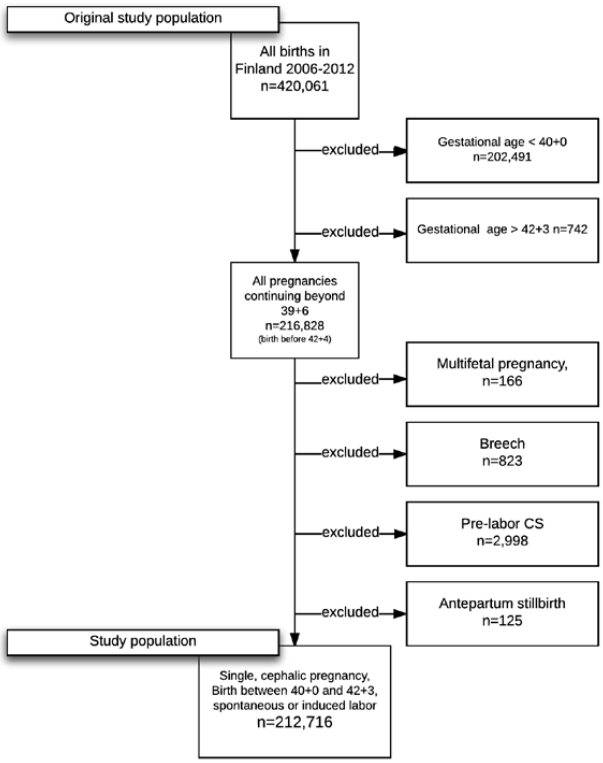


Figure 3 *Flow chart of the study population IV: The effect of labor induction in prolonged pregnancy*

4 DATA SOURCES

All studies presented in this thesis are based on registry data. The registry-based approach, a cost-effective and less resource-intensive method than the prospective study setup, is considered useful in quality assessment, especially for perinatal health where the registers are known to have very high specificity and also moderate to high sensitivity (190-192).

The main data source for Studies II-IV is the Finnish MBR. Study I, in turn, is based on data from all five Nordic MBRs (The Nordic Robson Research Collaboration, see below). The high quality data within the Nordic MBRs are unique and these registers are critical for large population-based studies on obstetrics. They also provide an excellent source for studies on quality of care offering clinical data that is comparable over time and between units and regions/countries (193-195).

4.1 THE FINNISH MEDICAL BIRTH REGISTER

The Finnish MBR was established in 1987 and is maintained by THL. Over the years, the register has gone through several reforms (1990, 1996, 2004, and latest 2017) to improve its reliability. There were no changes in the data gathering system in Studies II-IV, increasing the consistency of the research data.

The Finnish MBR gathers information for all births defined according to the WHO definition: Live born or stillborn infants born after the 22nd gestational week or weighing 500 g or more (196). The information is linked to information on all live births (Central Population Register) and on stillbirths and infant deaths (Cause-of-Death Register, Statistics Finland), resulting in virtually 100% coverage (30, 31).

The Finnish MBR data consist of information on pregnancy, birth, and neonates including both health-related (e.g. smoking, pregnancy complications, method of delivery, and neonatal outcomes) and socio-demographical (e.g. marital status, county of residence, place of birth) data. The extensive list of variables collected in the Finnish MBR is shown in the electronic form (Appendix 1); delivery units are required to provide this information for all births (as defined above). The GA is confirmed in maternal care units by a first trimester ultrasonography, as suggested by national guidelines (197).

In Finland, the collection of data is based on the Act on the National Institute for Health and Welfare (668/2008), the Act on the Statistical Service of the National Research and Development Centre for Welfare and Health (409/2001), and the Act on the National Personal Records Kept

under the Health Care System (556/1989) and the subsequent Decree (774/1989)(198). Data collection is obligatory and there is no opt-out possibility for the registered people.

4.2 OTHER NORDIC MEDICAL BIRTH REGISTERS

In all Nordic countries, there are three separate, partly overlapping registers / statistical systems gathering data on births. First, all live births are recorded in the Central Population Register, where by all born children are given a national identification code. This is essential for cross-linkage between different registers. Second, all stillbirths and deaths of live born children are registered in the Cause-of-Death Register, kept by statistical or health authorities. Third, all Nordic countries have MBRs for more detailed data collection of parturients, deliveries, and newborns (199). The MBRs in the Nordic countries are statutory and include data on all deliveries, covering all live births and stillbirths, which since 2008 have been universally determined according to the WHO definition (196).

The purpose of the national MBRs is data collection and maintenance with a secondary aim to develop and organize maternity care, obstetric services, and neonatal care through the information gathered. The coverage and accuracy of the national MBRs are close to 100% (30, 31, 200-202). The cross-linkage between population registries and different national care registries, like IVF, abortion, and national patient registers, further enhances the quality of the MBR data (203, 204).

Norway was the first to establish a separate MBR in 1967 followed by Denmark in 1968, Iceland in 1972, Sweden in 1973, and Finland in 1987. As in Finland, the Norwegian MBR is maintained by a public health institute (Institute of Public Health, Folkehelseinstitutet), whereas in Sweden the MBR is maintained by a national health authority (National Board of Health and Welfare, Socialstyrelsen). Also in Denmark, the national health authority maintained the MBR until 2015, when the maintenance was transferred to the newly established Danish Health Data Authority (Sundhedsdatastyrelsen). In Iceland, the corresponding organization is the University Hospital Landspítalinn in Reykjavík.

The information reported to the Nordic MBRs is based on compulsory notifications, with a largely similar content to the one used in Finland. All registers gather information on maternal socio-demographic background, previous pregnancies and deliveries, maternal diagnoses, care, and interventions during pregnancy and delivery, and information on newborn health, diagnoses, care and interventions (199). Mostly, parturient women can not refuse notification to the MBR, with a few exceptions, e.g. in Norway the notifications of maternal smoking habits, maternal use of alcohol and drugs, maternal occupation, and IVF treatment are voluntary.

4.3 THE NORDIC ROBSON RESEARCH COLLABORATION

The study on the CS trends in the Nordic countries (Study I) is based on an extensive dataset gathered by a joint effort of national researchers and statisticians. The dataset is based on data from national MBRs and includes detailed information on 5,716,725 deliveries between 1991 and 2011. The aim of the research collaboration to provide high-quality data on CS trends stratified by the Robson classification has directed the construction of the variables for the dataset. The dataset is under reconstruction and updating for further studies.

Despite the Nordic Robson dataset containing information from 1991 onwards, to ensure uniformity of the data, only the births after 1999 were included in the study presented in this thesis (I). For the data from 2000 onwards, there are very few cases with missing information on the essential variables regarding the Robson classification: parity and possible previous CS, mode of delivery, number of fetuses, onset of labor, and fetal presentation at onset of labor.

The study population was classified into Robson groups (for details, see Review of the Literature) respecting the hierarchy of the original definitions of the Robson groups (142).

5 STUDY SETUP AND STATISTICAL METHODS

5.1 STUDIED VARIABLES (I-IV)

To facilitate evaluation of different outcome indicators for quality and patient safety within obstetrics, the range of the variables was chosen to be as wide as possible. Considering the multidimensional nature of quality, we aimed at identifying all potential process indicators, and applied them along with the traditional outcome indicators. Structural elements are mostly included as background variables (e.g. country and size of the hospital). The variables used in the four studies of this thesis are shown in Table 6.

Table 5 *Variables used in Studies I-IV: background variables (B), background variables used for propensity score (PS), and outcome variables (O)*

Maternal variables	I	II	III	IV	Neonatal variables	I	II	III	IV	I
Age at birth, years	B	B	B	PS	Birth weight, g		B	O		
Previous births	B	B	B	PS	Date of birth ***				PS	
Pregestational BMI				PS	Time of birth***				PS	
Maternal smoking				PS	Perinatal death			O	O	
Previous CS	B	B		PS	Neonatal death			O		
Number of fetuses	B	B	B	PS	Early neonatal death			O	O	
Fetal presentation at the onset of labor	B	B			Stillbirths			O	O	
Infertility treatment*		B		PS	5 min Apgar score					
Maternity hospital		B	B	PS	Uc-pH < 6.95 / 7.0 / 7.05 / 7.10			O		
Gestational age at birth	B	B	B	PS	NICU			O	O	
Mode of delivery	O	B,O		O	Care in another hospital			O		
Epidural and/or spinal analgesia		O			Respirator treatment			O		
Labor induction	B	B		PS	Erb's paralysis			O	O	
Episiotomy		O			Fracture of clavicula			O	O	
Obstetric trauma		O		O	Birth trauma			O		
Maternal diagnosis during pregnancy**				PS	Prolonged hospitalization			O		

**) IVF, ICSI, FET, intrauterine insemination, ovulation induction*
***) 450 most typical diagnose codes on pregnant women according to the International Classification of Diseases, 10th revision (ICD-10) codes (see Appendix IV for details)*
****) Date of birth: day/month/year, Time of birth: hour/minute*

5.2 CESAREAN SECTION TRENDS IN THE NORDIC COUNTRIES (I)

The aim of the study was to analyze the trends in obstetric care by a stratified CS rate analysis using the Robson classification. This was done first by analyzing the overall CS trends for the whole population and for the individual countries, and second by studying contributing factors to the CS rate by analyzing individual Robson groups on four time periods.

5.2.1 Trend analysis with logistic regression

The 12-year study period analyzed in this study was divided into four three-year periods (T1 2000-2002, T2 2003-2005, T3 2006-2008 and T4 2009-2011). In the first phase, we analyzed the change in the total CS rate between T1 and T2-T4 using logistic regression adjusted for 1) maternal age, 2) maternal age and parity, and 3) maternal age and Robson groups.

5.2.2 Changes within the Robson groups

Subsequently, in order to illustrate the major contributors to the total CS rate, we merged the following Robson groups and provided their absolute contribution to the total CS rate (CS / all births), for each country and for each of the studied time periods:

- R1 and R2: single, cephalic, term nulliparous
- R3 and R4: single, cephalic, term multiparous (no previous CS)
- R5: previous CS and single, cephalic, term
- R6 and R7: breech presentation
- R8 and R10 (and R99): others (multiple, transverse fetal presentation and preterm pregnancies and uncategorized women)

Then, we studied further the individual Robson groups focusing on the groups that among the merged Robson groups contributed the most to the change in the CS rate (analyzed by calculating the crude change in the contribution between T1 and T4). Pearson correlation coefficient was used to estimate the correlation between the CS rate and the group size within individual Robson groups.

The statistical program used in all analysis was SAS software, Version 9.4 of the SAS System for Windows. SAS Institute Inc., Cary, NC, USA.

5.3 OBSTETRIC TRAUMA AND NEONATAL OUTCOME AND PROCESS MEASURES (II, III)

5.3.1 Comparison between hospital size categories

The outcomes were compared between the different hospital size categories as shown in Tables 7 and 8.

The arguments for including university clinics in the study on obstetric trauma but not in the study on adverse neonatal outcomes were based on different background risks and confounding factors. In the study on neonatal indicators, there are numerous potential confounding factors; to limit the factors related to high-risk pregnancies, the university clinics were not included in the main analyses. The background risk factors for obstetric trauma are more limited, and unlike for neonatal indicators, these factors are not clearly dependent on centralization. However, excluding the university clinics led to having only a single unit in the largest hospital size category which, in turn, led to the decision to use 3,000 (instead of 5,000) annual deliveries as a cut-off for large units in the study on neonatal outcomes. Also, we considered at cut-off below 1,000 (instead of 500) for small units to be more accurate due to these units being similar in regards to pediatric and anesthesiologist on-call arrangements, which play a critical role in neonatal care.

Table 6 *Hospital size categories used in the study on obstetric trauma (II).*

Hospital category	Number of units (university units)	Number of deliveries (% of total)		
		All units included		University clinics excluded
I <500	5 (0)	6,478 (2.2)	Small units	6,478 (3.4)
II 500–999	10 (0)	34,887 (12)		34,887 (18)
III 1000–1999	8 (0)	56,015 (19)		56,015 (30)
IV 2000–2999	4 (1)	43,161 (15)	Reference group	30,981 (16)
V 3000–4999	4 (2)	73,618 (25)		32,338 (20)
VI >5000	3 (2)	80,566 (27)	Large units	28,713 (15)
Total	34 (5)	294,725		189,432

Table 7 *Hospital size categories used in the study on neonatal outcome and process measures (III)*

Hospital category	Number of units (university clinics)	Number of deliveries (% of total)		
		All units included	University clinics excluded	
I <500	5 (0)	6,232 (2.3)	6,232 (3.5)	Small units
II 500–999	10 (0)	33,634 (12)	33,634 (19)	
III 1000–1999	8 (0)	53,092 (19)	53 092 (30)	Reference group
IV 2000–2999	4 (1)	40,483 (15)	29,254 (16)	
V 3000–4999	4 (2)	68,949 (25)	30,791 (17)	Large units
VI >5000	3 (2)	73,897 (26)	27,365 (15)	
Total	34 (5)	276,287	180,368	

In 2006, at the beginning of the study period, there were 34 birth units in Finland but by January 2017, the number had decreased to 26 (two units closed during the study period, between 2006 and2010). In 2017, only six units had a volume of less than 1,000 annual deliveries, of which one less than 500 (29).

5.3.2 Maternal indicators (II)

In the study on obstetric trauma, the primary outcome variable was obstetric trauma, which was analyzed for the total population and for the sub-populations shown in Figure 2 (instrumental and non-instrumental vaginal delivery) for each hospital size category. The secondary outcomes were the following direct process measures analyzed for the total population: the mode of delivery (CS and instrumental and non-instrumental vaginal delivery), the use of episiotomy and epidural and/or spinal analgesia.

5.3.3 Neonatal indicators (III)

In the first phase, we analyzed variables chosen based on previous publications (24, 25, 111, 205, 206), on the availability of the registry data needed for the indicators, and their feasibility in Finland (Table 6). For most of the studied variables, the reporting levels were nearly 100%.

Of the outcome indicators, the most problematic one was uc-pH, where the reporting levels varied greatly between the delivery units - from 0 to 96% of births were reported with an uc-pH value (average 76%, 11 units reported less than 50%) decreasing the validity of the results. The indicator was still included in the analysis, but the units reporting less than 10% (n=6) were

excluded. After the exclusion, the average reporting rates were 67% in the small, 77% in the reference units and 92% in the large units, and the study population was reduced from 189,433 to 154,018.

To test the validity of birth trauma in the Finnish setting we adjusted the definition of the indicator to the International Classification of Diseases, 10th revision (ICD-10) codes (1, 207). In addition, we used Erb's paralysis and clavicular fracture as individual birth trauma measures.

The potential process measures proved more complicated to use; admittance to NICU was omitted because with this specific parameter, there are known differences in terminology and reporting policies between hospital districts, which would have confounded the results excessively. Instead, we studied the possibility to use "transfer to another unit", but unfortunately this, too, proved futile due to variation in structural factors like the location and terminology for a neonatal unit.

After these preliminary analyses, the following outcome and process measures were chosen as outcome variables for the study:

- Perinatal mortality: number of stillbirths and deaths in the first week of life per 1,000 live births
- Neonatal mortality: number of deaths during the first 28 days of per 1,000 live births
- Early neonatal mortality (ENM): number of deaths in the first week of life per 1,000 live births
- Stillbirths (both antepartum and during labor)
- 5-min Apgar <4 and <7 (208)
- Umbilical cord arterial pH < 6.95, 7.00, 7.05, and 7.10
- Erb's paralysis: Discharges with ICD-10 codes P14
- Fracture of clavicula: Discharges with ICD-10 codes P13.4
- Birth trauma: Discharges with ICD-10 codes (P10-15) for birth trauma and (P52) for intraventricular nontraumatic hemorrhage, including all deliveries with a newborn weighing more than 2,000g (modified from AHRQ definition) (1)
- Respirator treatment
- Prolonged hospitalization of a neonate: ≥ 7 days after birth
- Proportion of postterm deliveries (GA ≥ 42^{+0})

5.3.4 Statistical methods

To analyze the differences in adverse outcome rates between the hospital size categories, we used logistic regression analysis adjusted for maternal age and parity. For the study on obstetric trauma, we also analyzed the risk between different Robson groups (regardless of the hospital category): 1) the risk in in induced labor (Robson 2a vs. 4a) 2) and in labor with a spontaneous onset in nulliparous versus multiparous deliveries (Robson group 1 vs. 3), and 2) the

risk in spontaneous versus induced labor in nulliparous (1 vs. 2a) and in multiparous deliveries (3 vs. 4a), separately.

The statistical data were managed with SPSS for Windows 17.0 (SPSS Inc, Chicago, IL, USA).

5.4 ANALYSIS OF THE EFFECT OF LABOR INDUCTION IN PROLONGED PREGNANCY (IV)

5.4.1 Studied outcomes

The maternal and neonatal outcome variables used in the study were: emergency CS (vs. all vaginal deliveries), operative delivery (both emergency CS and instrumental vaginal delivery vs. spontaneous vaginal delivery), obstetric trauma i.e. 3rd or 4th degree perineal rupture, 5-min Apgar <7, respirator use on neonate, meconium aspiration syndrome, prolonged hospitalization of neonate (≥7 days), stillbirth during labor and perinatal mortality.

5.4.2 Case and control

Our study group consisted of all induced births within the study population. To analyze the outcomes on different gestational ages, we divided the study groups based on the GA at delivery into five three-day groups: I) 40^{+0} to 40^{+2}, II) 40^{+3} to 40^{+5}, III) 40^{+6} to $41^{+1,}$ IV) 41^{+2} 41^{+4}, and V) 41+5 to 42+0 (Table 9).

Table 8 *Gestational age periods and the number of births, induced births (case), expectantly managed pregnancies and the propensity score (PS) matched controls.*

	Gestational age period	Total births	Induced births	Expectantly managed pregnancies	PS matched* controls	Total study population
I	40^{+0} - 40^{+2}	51,364	6,882	205,834	6,882	212,716
II	40^{+3} - 40^{+5}	49,077	5,543	155,809	5,543	161,352
III	40^{+6} - 41^{+1}	42,065	5,115	107,160	5,115	112,275
IV	41^{+2} - 41^{+4}	32,889	5,581	64, 629	5,581	70,210
IV	41^{+5} - 42^{+0}	26,369	10,167	27,154	10,167	37,321

*) Propensity score matching was performed using independent variables: parity; maternal age; body mass index (BMI); smoking; previous CS; infertility treatment; birth unit, year, month and weekday of birth along with 450 ICD-10 diagnosis codes grouped into 12 categories.

For each of these study groups (induced labor), we formed a respective control group (expectant management) that consisted of all deliveries after the study period as well as all spontaneous deliveries during the study period. The latter was done in order to avoid overestimating the risk for CS in the study group as suggested previously (174, 177)

5.4.3 Propensity score method

> *"The propensity score is a one-dimensional variable that summarizes the multidimensional pretreatment covariates"*
>
> *Kurth et al (209).*

In our study setup, propensity score (PS) represents the probability of labor induction relative to expectant management given the used covariates (see below). Effectively, matching the cases (labor induction) with controls (expectant management) using covariate balanced PS simulates RCT in a cohort study setup: PS matching allows causal interpretations because matched pairs conceptually represent the same patient under different treatment scenarios reducing the confounding by indication. In addition, a far greater array of independent variables can be used in constructing the PS than in traditional regression models enabling control over a vast set of possible confounding factors.

The variables used for covariate balanced PS (210) matching in our study included the following: parity; maternal age; body mass index (BMI); smoking; previous section; infertility treatment; labor unit, year, month, and weekday of birth along with 450 most typical diagnose codes on pregnant women (according to ICD-10(207)). Supplementary data in the publication IV contains detailed information about the variables used.

We matched each case (labor induction) 1:1 with an expectantly managed control and performed PS matched Poisson regression to form a relative risk (RR) for each of the studied outcomes. In addition, we performed logistic regression to obtain odd ratios (ORs). The Poisson regression and logistic regression analysis were performed unadjusted, adjusted for the independent variables used in the PS construction, and with PS matching. Since with common outcomes (prevalence 10% or more), like some used in our study, ORs produced by logistic regression do not approximate RRs well, we report only the results of PS matched Poisson regression which we regard as the best suited for the study purposes (211).

The analyses were carried out by R language (212).

6 ETHICS

6.1 CESAREAN SECTION TRENDS IN THE NORDIC COUNTRIES (I)

The Danish Data Protection Agency governed the Danish participation (reference NOH-2016-006, med I-Suite no. 04548). The Regional Committee for Medical and Health Research Ethics, South-East C (REK 2010/3256) assessed the Norwegian participation. The Directorate of Health in Iceland, THL in Finland, and the Swedish National Board of Health and Welfare, which maintain or supervise the national MBRs, gave their authorizations for the use of the aggregated anonymized registry data analyzed in this study, and therefore no ethical approval was needed.

6.2 MATERNAL AND NEONATAL OUTCOMES IN DIFFERENT-SIZED BIRTH UNITS AND IN LABOR INDUCTION (II-IV)

THL National Institute for Health and Welfare is authorized to disclose data in the Finnish MBR to researchers for scientific research purposes after consulting the Data Protection Ombudsman. As required by national data protection legislation, the research for this thesis obtained permission from THL to use the anonymized data.

The data subjects have no right of access to and no right to rectify the data entered into the register because the Medical Birth Register is a statutory statistical and research register and the personal data stored in it are not used in decision-making or care concerning the data subjects. Therefore, no informed consent was needed from the study population.

The data stored in the Medical Birth Register are confidential under section 4 of the Act on National Personal Data Registers Kept under the Health Care System (556/1989) and the data used in the thesis research were maintained in accordance with the THL data security guidelines.

For the study on the effect of labor induction an optional ethical approval was sought from and granted by the Research Ethics Committee of the Hospital District of Helsinki and Uusimaa (IRB00003181).

RESULTS AND INTERPRETATION

7 CESAREAN SECTION TRENDS IN THE NORDIC COUNTRIES - STRATIFIED ANALYSIS WITH THE ROBSON CLASSIFICATION (I)

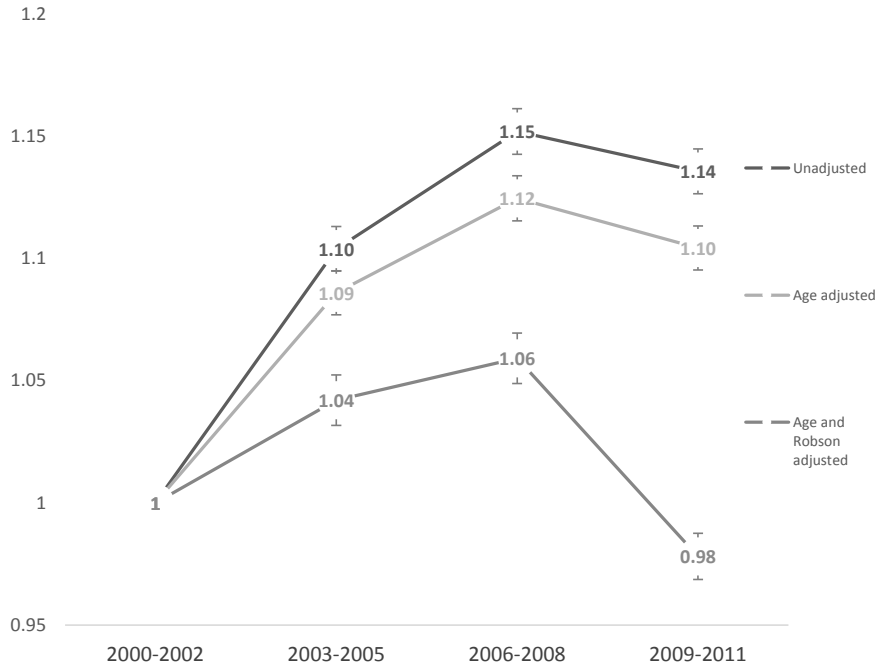
7.1 STABILIZING CESAREAN SECTION TREND

Between 2000-02 and 2009-11, for all of the Nordic countries combined the CS rate increased, but most of the increase took place at the beginning of the millennium while towards the end of the study period the rate stabilized or even decreased. Combined, the total CS rate during the studied time periods was: 15.7% in 2000-02, 17.1% in 2003-05, 17.7% in 2006-08, and 17-5% in 2009-11.

In Figure 4, the trends in the CS rates are shown as ORs for three-year periods using the first period as a reference. The increase is partially explained by changes in the parturient population's age and parity. When adjusting by Robson groups and parity, the bell-shaped trend is flattened and the probability for CS is lower at the end than at the beginning of the study period. This means that the variables used in the Robson classification, both the changes in the parturient population and in the obstetric processes, like labor induction and pre-labor CS, explain the increase in the CS rates between 2000 and 2011. To understand the actual changes that have taken place, the population must be studied by the individual Robson groups; results of the Nordic data are presented in the following chapters.

There are variable national CS rates behind the bell shaped overall CS trend. The total CS rate increased in Denmark (age-adjusted OR 1.28, 95% CI 1.25-1.30), Norway (OR 1.08, 1.07-1.10), and Sweden (OR 1.14, 1.12-1.16) over the total study period. In Finland, there was a slight decrease (OR 0.97, 0.96-0.99) and in Iceland a substantial decrease (OR 0.82, 0.77-0.88) in the CS rate.

Between 1990 and 2014, most of the global CS rate increase took place in the least and less developed regions of the world (26) and reassuringly, there are reports also from other developed countries , e.g. in Belgium, France, Italy, UK and USA, of stabilizing CS rates towards 2010 (25, 213, 214).



Finland	16.5	16.5	16.5	16.2
Denmark	16.4	19.5	20.5	20.7
Iceland	17.5	16.8	17.0	15.3
Norway	14.4	15.5	16.3	16.5
Sweden	15.5	16.8	17.5	17.1

Figure 4 *Trends in the CS rates in all Nordic countries 2000-2011 (figure) and the national CS rates (table) on each time period. Odd ratio (OR) for CS rate for different time periods using 2000 to 2002 as a reference.*

7.2 ABSOLUTE CONTRIBUTIONS OF THE ROBSON GROUPS

7.2.1 Significance of nulliparous women and women with a previous Cesarean section

The absolute contribution of a Robson group is the number of Cesarean sections in that group per all births – therefore, the total CS rate is the sum of absolute contributions of different Robson groups. In our study, the Robson groups were merged into clinically relevant sets described in the Methods section. The contributions of these merged Robson groups are shown along

with the total CS rate for each country and each time period in Figure 5. In this way, the major changes contributing to changes in a total CS rate can be communicated in one illustrative figure offering a comprehensible look at the "big picture".

In all countries and time periods, most CSs (47-64%) were performed on term singleton cephalic nulliparous women (R1-R2) and on multiparous women with a single, term, cephalic pregnancy but with a previous CS (R5). These groups are known to be the major contributors to the total CS rate (147, 215) and similar relative contributions have been reported from other populations, too: 60% in France in 1995-2010 (213), 50% in Peru in 2008-2010 (216), and 63% in a large multicountry reference population from outside Europe and USA (144).

The contributions of R1-R2 and R5 changed significantly in all countries and explained nearly all of the increase seen in the total CS rate in Denmark, Norway and Sweden. This is in accordance with previous reports, too (213, 217, 218). In Iceland, where the total CS rate decreased, the absolute contribution of R1-R2 decreased substantially explaining about 25% of the decrease in the total CS rate. This makes Iceland a unique example (discussed below).

7.2.2 Relative size and Cesarean Section rate within a Robson group constitute the absolute contribution

The absolute contribution of a Robson group (CS in a group / all births) is essentially the CS rate within a group (CS in a group / births in a group) multiplied by the group's relative size (births in a group / all births). This highlights how opposite changes in these two parameters may result in a steady contribution from a Robson group. This, in turn, may cause misinterpretations in a trend analysis relying on changes in absolute contributions: e.g. if the proportion of nulliparous women decreases (relative size of R1-R2) but the CS rate among nulliparous women increases (CS rate of R1-R2) one may draw the false conclusion that the obstetric care of nulliparous women has not changed. Therefore, studying the underlying changes in the relative size and the CS rate of each Robson group are essential before drawing conclusions on findings based on the absolute contribution of a Robson group (or a merged Robson group).

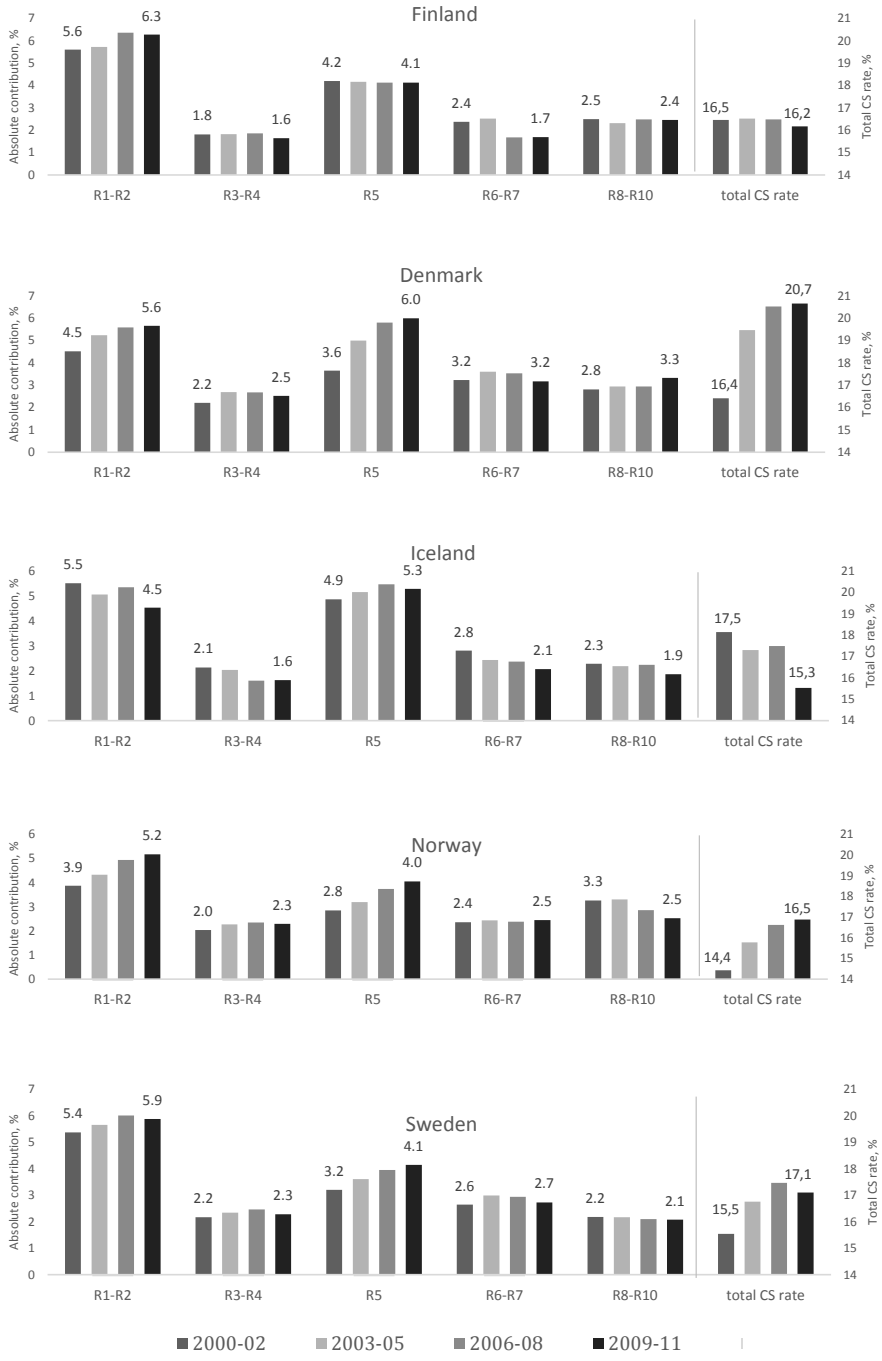


Figure 5 *Total Cesarean section (CS) rate and the absolute contributions of the merged Robson groups to the total CS rate by country and time period studied.*

7.3 CHANGES WITHIN INDIVIDUAL ROBSON GROUPS

7.3.1 Robson groups 1 and 2 (R1-R2): term singleton cephalic nullipara

The clinically most relevant finding was the increased contribution of R1-R2.

A significant finding in the analysis of the individual Robson groups R1 (spontaneous onset of labor), R2a (induced labor) and R2b (pre-labor CS) was the steadily increasing contribution from R2a in all countries. This was explained by the increase in the relative size of the group from 4-5.6% to 6.2-8.7% - i.e. at the end of the study, about 50% more nulliparous women with a single, term cephalic pregnancy had their labor induced.

Changes in parity (increased proportion of nulliparous women) affected the increased contribution from R1-R2 significantly only in Finland and Norway accounting for about 30% of the increase. Interestingly, in Iceland the proportion of nulliparous women decreased and explained about 16% of the decrease in the contribution from R1-R2.

The relative sizes of R1, R2a and R2b within R1-R2 are shown in Table 2 in the original publication (see Appendix). These parameters are exclusively dependent on obstetric practices and changes in them reflect changes in obstetric care. It is important to acknowledge that the increase in the induction rates on nulliparous women did not directly translate to increased overall contribution from R1-R2 i.e. does not (completely) explain the increase in the CS rate in the low-risk nulliparous group. Iceland providess an illustrative example of how there is no direct correlation between the induction rate and the overall CS rate; in Iceland, at the end of the study, the CS rate for R1-R2 was the lowest, while the relative size of R2a was the highest.

Despite the increase in the contribution from R1-R2, the CS rate among this group remained very low in all Nordic countries, between 12.6% and 16.8% at the end of the study. In an international multicenter study, the CS rate for term singleton cephalic nullipara was reported to lie between 14.6% and 31.2% in individual institutions (219).

7.3.2 Robson groups 3 and 4 (R3-R4): term singleton cephalic Multipara (no previous CS)

The contribution from R3-R4 remained steady.

Induction rates increased also in multiparous (R3-R4) women: the relative size of R4a increased from 4.0-6.9% to 5.1-9.3%. The contribution of R3-R4 to the total CS rate increased slightly in Denmark, Norway and Sweden but these increases contributed less than 10% of the increase in the total CS rate. In Iceland, where the contribution decreased, the impact to the total CS rate decrease was only half of that of nulliparous women (R1-R2).

Merging the contributions of R1-R4 offers relevant additional information; these Robson groups consist solely of women without a previous CS i.e. they form the most significant population at risk for the first CS. An essential element of obstetric quality improvement is to reduce the risk for the first CS and therefore, it is of high clinical relevance to look at these parturient groups together (220).

Based on our results and previous studies, R1-R4 contribute nearly half of the total Cesareans in any population; in our data, between 40 and 49% of the cesarean sections were performed on women in these groups (Denmark and Iceland 40%, Finland 49%, Norway 45% and Sweden 48%) compared with 47% (total CS rate 15.6%) in the Netherlands (221), and 43% (total CS rate 20.5%) in France(213). In the multinational reference population formed by Souza et al, the respective proportion was 49% and the total CS rate 18.5% (144). However, the association between the contribution of R1-R4 and total CS rate seems weak. More sensitive trend analysis could potentially reveal causality between an increasing or decreasing CS trend and the relative contribution from these low-risk Robson groups.

7.3.3 Robson group 5 (R5): term singleton cephalic Multipara with a previous CS

R5 was the other major contributor to an increasing overall CS rate (in addition to R1-R2).

The absolute contribution of R5 increased in Denmark (3.6% to 6.0%), Iceland (4.9% to 5.3%), Norway (2.9% to 4.1%) and Sweden (3.2% to 4.1%). In Finland, the contribution from R5 remained stable (4.2% to 4.1%) due to opposite changes in group size and CS rate. This offers an example of how by studying the absolute contributions only, the underlying changes in obstetric processes may go unnoticed.

As expected with increasing total CS rates, the relative size of R5 increased slightly in all Nordic countries (from 6.5-9.2% to 7.8-10.3%) but

remained lower than in most other populations (144, 215). Exceptions in previous reports were the Netherlands, with a very low total CS rate (15.6%), where the relative size of R5 was only 6% (221), and, surprisingly, Peru, where the total CS rate has been over 24% for over a decade, but the relative size of R5 was only 8.4% (216). In our study, Sweden and Norway had a lower relative size of R5 than the other Nordic countries (8.2% and 7.9% vs, 9.7-10.3% between 2009 and 2011). For Norway, this is likely due to the historically low CS rate while for Sweden, the finding might be explained by the larger overall proportion of nullipara. The relative size of R5 and the overall CS rate, and its trends, are no doubt correlated, but in light of our results the correlation is not as straightforward as one might expect.

Approximately half of the increase in the contribution of R5 in Denmark, Norway and Sweden was explained by the increase in group size (i.e. increased number of women with a previous CS) – rest was due to the increased CS rates for R5. The CS rate for R5 increased in Denmark (46% to 59%), Norway (42% to 50%) and Sweden (49% to 53%) but decreased in Finland (46% to 43%) and Iceland (56% to 51%). Compared with international figures, all Nordic countries had very low CS rates for R5: in previous studies the CS rates have ranged between 61% and 80% (147, 213, 216, 217, 219).

Due to the small relative size and significantly lower CS rate for R5 than in most other populations, the absolute contributions of R5 remained low despite the increases in all Nordic countries - only about half of the average contribution of R5 reported for highly developed countries (147).

7.3.4 Robson groups 6 and 7 (R6-R7): All Nulliparous (R6) and multiparous (R7) women with breech presentation

The Nordic countries differed significantly with regard to CS rate in breech deliveries

In Finland and Iceland, with decreasing total CS rates, the contribution of R6-R7 (all breech) decreased significantly. It was mostly this change that compensated the increased contribution of R1-R2 and resulted in the slightly decreasing total CS rate in Finland.

Interestingly, the study on the relative sizes and CS rates among this merged Robson group showed that the Nordic countries differed even more than was expected by the absolute contributions. These figures are seen in the supplementary material of the original publication. The CS rates within R6-R7 showed that Finland and Norway are in a different range than the other countries (CS rate in R6-R7 68% in Finland and 71% inNorway for the last time period vs. 89-91% in the other Nordic countries). The Finnish and Norwegian term breech CS rates are significantly lower than in other

reported populations; even in the Netherlands, where the total CS rate was very low (15.6%), the CS rate for R6-7 was 79%(221)

Although R6-R7 do not contribute substantially to the overall CS rate, the wide variance in the management of this obstetric subpopulation is an interesting finding – especially considering the secondary aims of this thesis in facilitating safe, effective and equitable care by contributing to best practice protocols. The term breech trial, a multicenter study comparing planned CS and vaginal birth for breech delivery (164), has had a significant impact on the management of breech deliveries and has substantially increased the CS rate among R6-R7 globally. The term breech trial and its massive impact have been questioned; Glezerman stated in 2006: "Rarely in medical history have the results of a single research study so profoundly and so ubiquitously changed practice"(165). A registry-based study in the Nordic countries focusing on the maternal and neonatal outcomes in R6 and R7 could contribute to this matter.

7.3.5 Robson group 8: All multiple pregnancies

There is a high proportion of multiple pregnancies in Denmark.

The contribution from women with multiple pregnancies to the total CS rate is not substantial and therefore, detailed analysis of R8 was beyond the scope of this study. However, we did find out that the Nordic countries were rather heterogeneous with regard to women with multiple pregnancies. The CS rate with R8 varied between countries: Denmark 57%, Finland 49%, Iceland 41%, Norway 45%, and Sweden 54%. The relative size of R8 was especially high in Denmark (2.3%) likely reflecting differing IVF policy from the Nordic countries, which showed more consistent number of multiple pregnancies (Finland 1.5%, Iceland 1.4%, Norway 1.7% and Sweden 1.4%).

In the Netherlands, with a very low total CS rate (15.6%) the relative size of R8 was 1.8% and the CS rate 43.1%. In the multinational reference population, the respective numbers were 0.9% and 58% (144), and in a tertiary center in Singapore 1.7% and 70% (221).

7.3.6 Robson groups R9, All abnormal lies, and R99, uncategorized women (missing data)

R9 and R99 enable quality control

The relative size of R99 (women not classified due to missing data), along with the relative size and the CS rate within R9 (all singleton pregnancies where the fetus is in transverse presentation at onset of birth) are considered

valuable parameters for auditing the quality of the data (149, 222). In our study, the relative size of R99 was very low compared with previous studies (0.-1.4% vs. 2-3.6%) (144, 213, 216). Also the relative size of R9 was within the suggested limits 0.2-0.6% in all countries except for Sweden (0.1%) (222). In turn, the CS rate in R9 was not 100%, as it should be, in all countries (Denmark 95%, Finland 68%, and Sweden 91%), indicating that the classification process was not completely flawless.

7.4 COMPREHENSIVE REDUCTION IN THE CESAREAN SECTION RATES FOLLOWING A MONITORING PROGRAM

Iceland implemented a CS monitoring program based on the Robson classification already in the year 2000 (145). Within the study period from 2000 to 2011, the total CS rate in Iceland decreased from the highest to the lowest among the Nordic countries. This sets a clear example of how significant positive results can be achieved by benchmarking and implementing a monitoring program – an association that needs further validation. In a small population like that in Iceland (approximately 300,000 during the study period), the changes are likely to be seen more rapidly, but their situation suggests that also other, larger populations could experience similar results

7.5 STRENGTHS AND LIMITATIONS OF THE STUDY

Detailed data on over 3 million births strengthen the validity of the findings. The study population of 3.4 million births is larger than in nearly all previous Robson studies (147, 213, 216, 218, 223, 224). More important than the absolute size of the population, is that it covers virtually 100% of all births during the study period, which greatly enhances the validity of the data. An additional strength is that the analysis is provided from countries with sustainably low total CS rates – especially important in the era when over just two decades, the CS rates have approximately doubled in the developed world (from 14.5% to 27.2% between 1990 and 2014) (26, 137, 225). The figures provided in the study offer usable benchmarks; due to the stratification with the Robson classification, they are comparable over time and between populations.

The Robson classification incorporates some of the background variables of a parturient population while other essential elements, such as maternal age and BMI, are excluded. Morbidities like gestational diabetes, pre-eclampsia or fetal growth restriction are not taken into account. Given that the Nordic population is known to be rather homogeneous in many of these aspects, the Robson analysis is likely to provide reliable stratification for

comparisons between countries in our study. Comparisons over time, in turn, may be confounded by the increased maternal age and BMI, as well as the more frequently seen morbidities like gestational diabetes (225, 226).

CS rate analysis with the Robson classification warrants caution. The classification system may appear straightforward, but the interpretation of the results requires comprehension of the limitations (above) and of the interdependence of the Robson groups. The changes in the contributions of different Robson groups, which together constitute the total CS rate, can be communicated in a somewhat simple manner, but the information is short and possibly misleading if the analysis is not completed by studying the CS rates and relative sizes of each individual group. If this is not done, the possible underlying changes in obstetric care will be disregarded (e.g. stable absolute contribution, but substantial changes in the CS rate and size for R5 in Finland).

Subsequently, if the interdependency of the groups is not acknowledged, conclusions drawn may be completely opposite to the truth; e.g. a decrease in the CS rate within R2a (induced labor) may have a negative impact on the total CS rate when it results from a significantly decreased number of nulliparous women giving birth spontaneously (R1).

8 DIFFERENCES IN OBSTETRIC AND PERINATAL CARE BETWEEN BIRTH UNITS – ANALYSIS WITH POTENTIAL MEASURES FOR QUALITY ASSESSMENT (II, III)

8.1 OBSTETRIC TRAUMA

8.1.1 Inter-unit variation in obstetric trauma rates

Over the whole study period (2006-2010), the national obstetric trauma rate was 0.95% in all vaginal deliveries in Finland (0.67% and 3.43% in non-instrumental and instrumental vaginal deliveries, respectively). When analyzed for all 34 birth units, the average obstetric trauma rate in each unit was 0.86% with a range from 0 to 3.1%, mean (standard deviation, SD) 0.74 (0.63). The respective means for non-instrumental and instrumental vaginal deliveries were 0.62 (0.48) and 3.2 (2.4). The obstetric trauma rates are shown by the size of the unit in a scatter plot in Figure 6. The figure displays the great inter-unit variation in trauma rates, even within hospital size categories.

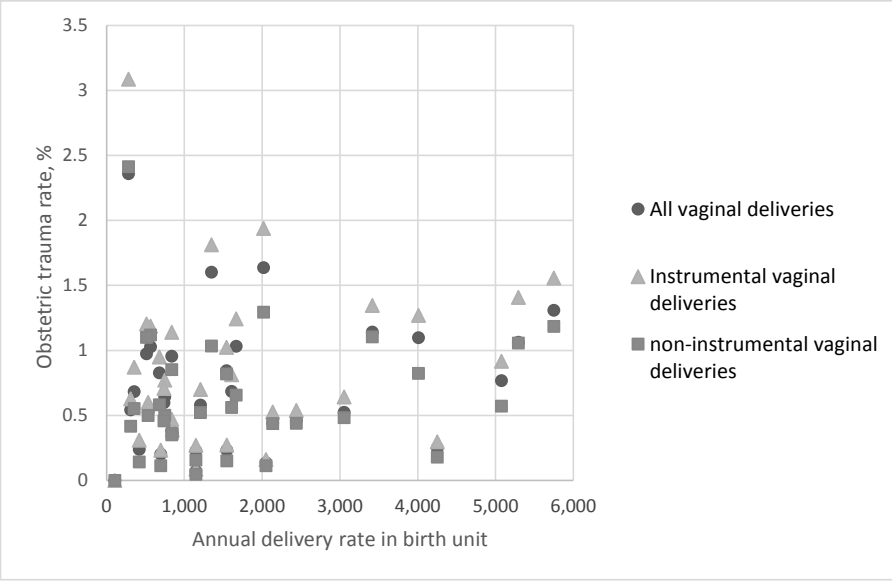


Figure 6 *Obstetric trauma rate (%) by the size of the birth units in Finland. Polynomial trendline for instrumental and non-instrumental vaginal deliveries.*

The obstetric trauma rates are shown by hospital size category in Figure 7. The polynomial trendline with high R^2 value suggests that there is u-shape association between the hospital size category and the risk for obstetric trauma. The mid-sized units (500 to 4,999 annual deliveries) these were merged and used as a reference (mean obstetric trauma rates 0.8%, SD 0.06). When adjusted by maternal age and parity, the risk was about 33% higher in the largest and 46% higher in the smallest units than in the mid-sized units. In the largest units, the risk for obstetric trauma was elevated in both instrumental (OR 1.26, 95% CI 1.10-1.45) and non-instrumental vaginal deliveries (OR 1.64, 95% CI 1.48-1.83). In these sub-classes, the risk for obstetric trauma in the smallest units was not statistically different from the reference group. However, it is important to acknowledge the great inter-unit variation within hospital size-categories as seen in Figure 6.

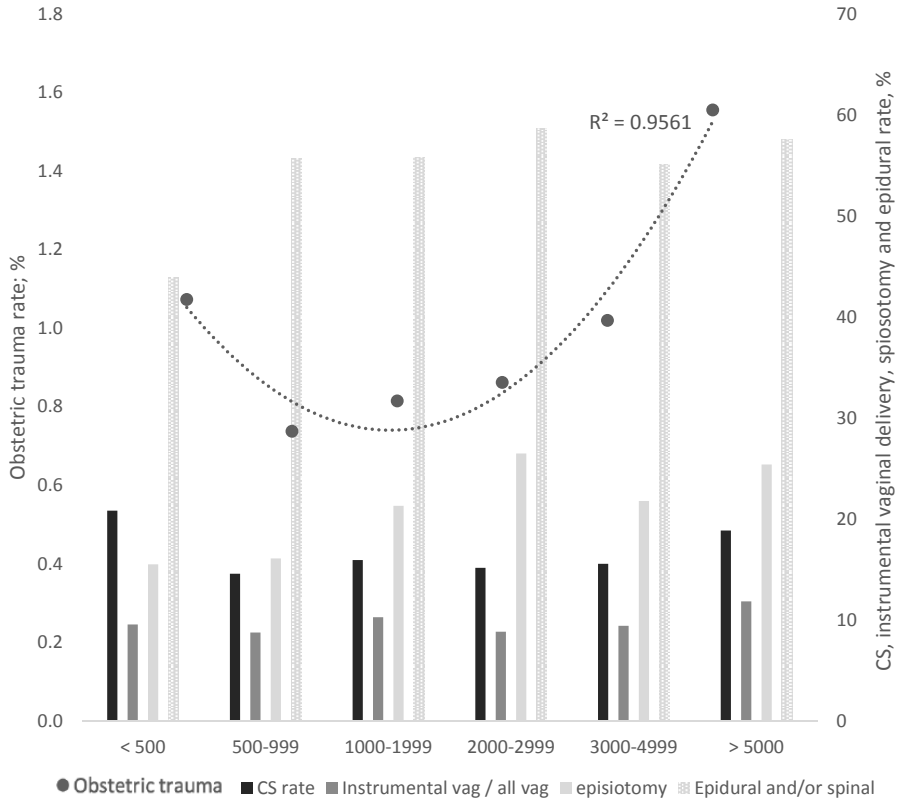


Figure 7 *Obstetric trauma rate and the rate of Cesarean sections (CS), instrumental vaginal deliveries, episiotomies and epidural and/or spinal analgesia in different-sized delivery units. Polynomial trendline for the obstetric trauma rates per unit size categories.*

When analyzed for the Robson groups 1-5, the differences between hospital size-categories remained. Interestingly, in some groups the differences were highly pronounced; in the small units, the risk was more than twofold among women with a previous CS (R5, OR 2.20, 95% CI 1.14-4.22) and among multiparous women with a spontaneous onset of labor (R3, OR 2.90, 95% CI 1.68-5.02, age-adjusted) compared with mid-sized units. In the large units, the risk increase in R1-R5 was slightly higher than when analyzed for the total population, 34 % to 47% (R1, OR 1.44, 1.28-1.61 and R2a, OR 1.34, 95% CI 1.06-1.63 and R3, OR 1.37, 95% CI 1.06-1.77 and R5, OR 1.4, 95% CI 1.14-1.90, R4a was not statistically significantly elevated OR 0.97, 95% CI 0.54-1.74).

According to our results, a high birth weight does not explain the inter-unit differences in the obstetric trauma rate. Previously, that has been associated to an increased risk for obstetric trauma (227-229). In our study, the neonate weighed > 4,000g in 2.8% of the births.

8.1.2 Inter-unit variation in obstetric trauma rates is a patient safety issue

The higher rate of potentially preventable adverse outcomes in the largest and smallest units is a patient safety concern. We believe the different incidences are more strongly related to care provider (true differences in care and possible reporting bias) than to obstetric population (case-mix). The increasing awareness and improved identification of the trauma may have increased the reporting in some units compared with others (reporting bias (230)) but also, there are likely to be true differences in obstetric care, more specifically in the management of the second stage of labor, e.g. in manual support of the perineum and in use of episiotomy. Both the obstetric management and the diagnostics and reporting are important areas for quality improvement; differences in outcome rates indicate there is room for benchmarking and education.

In a clinical intervention program in Norway, midwives were taught to slow down the delivery of the infant's head and to instruct the parturient not to push actively. This resulted in a significant decline in the incidence of the trauma and provides an example of an exceptional quality indicator (121, 231). First, the increasing trend in obstetric trauma pointed out an area ripe for improvement. Second, benchmarking to Finland revealed the rates were significantly higher in Norway, and finally, a quality improvement program, also called the Finnish intervention, was implemented and successfully executed (232).

8.2 OBSTETRIC PROCESS MEASURES

8.2.1 U-shape association between CS rate and volume of a birth unit.

During the study period (2006-2010) the CS rate in Finland was 16.4%. Per unit, CS rate was 16.2 (SD 3.3).

In the analysis comparing very small birth units (less than 500 annual deliveries) with mid-sized units, we found 35% higher risk for CS in the smallest units (RR 1.35, 95% CI 1.29-1.42, adjusted for maternal age and parity) and 22% higher risk in the largest units (RR 1.22, 95% CI 1.20-1.24) (Figure 7).

Subsequently, we studied the differences in the CS rates using the Robson classification. Using the same hospital categories, but limiting the analysis to R1-R2 (term singleton cephalic nullipara), we found that the CS rate was over 50% higher in the smallest and in the largest units (RR 1.82, 95% CI 1.68-1.96 and OR 1.51, 95% CI 1.47-1.56, respectively) than in the mid-sized units. Hence, the differences between the different-sized delivery units were much more pronounced (82% and 51% increase) among term singleton cephalic nullipara than when studying the overall CS rates.

With the university clinics analyzed separately, the study revealed the university clinics had a lower CS rate for R1-R2 (RR 0.86, 95% CI 0.82-0.90) than the mid-sized units. This was an unexpected finding especially considering the limitations of the Robson classification with regard to neonatal and maternal morbidity (e.g. intrauterine growth restriction and gestational diabetes mellitus); despite the centralization of high-risk pregnancies to the university clinics, the CS rate was lower in these units than in the mid-sized units catering mainly to obstetric population with low morbidity. This highlights how the CS rate is more dependent on the obstetric processes than on the background population.

The instrumental vaginal delivery rates were increased only in the largest units (absolute instrumental vaginal delivery rate 9.6%, RR 1.21, 95% CI 1.18-1.24). The corresponding rates were 7.6% in the smallest units and 7.7% in the mid-sized units. The results were similar when analyzing the proportion of instrumental deliveries of all vaginal deliveries. The instrumental delivery rate is shown for all vaginal deliveries by each hospital size category in Figure 7 (instrumental vaginal deliveries / all vaginal deliveries).

8.2.2 High Cesarean section rate does not decrease the obstetric trauma rate

Figure 7 presents a correlation between high CS rate and a high obstetric trauma rate correlate by the hospital size categories. Despite the very low correlation (correlation coefficient, CC 0.11 when analyzed for hospital size categories, CC 0.20 when analyzed for individual units), there is still a positive correlation between these two rates. This indicates that the obstetric trauma cannot be reduced by simply having less women giving birth vaginally. To our knowledge, this has not been previously reported.

As expected, there was a positive correlation between instrumental vaginal delivery rate and obstetric trauma rate (CC 0.81 studied for hospital size categories) – instrumental vaginal delivery being one of the most important risk factors for obstetric trauma (118, 233). However, the individual units differed markedly in how the instrumental vaginal delivery rate and obstetric trauma rate correlated (CC 0.21). In fact, in the smallest units (<500 annual deliveries), the correlation was found to be negative (CC -0.45). These findings indicate that instrumental vaginal delivery practices vary significantly between birth units and there may be underlying differences in e.g. physicians' experience/training in vacuum extraction and in how frequently adjuvant episiotomy is used.

8.2.3 Episiotomy and epidural analgesia more frequently used in the large units – no association with birth trauma

The episiotomy rate was increased in the large units and decreased in the small units relative to the mid-sized units. Similarly, the epidural and/or spinal analgesia was used more frequently in the largest and less frequently in the smallest units (Figure 7).

Unfortunately, we were unable to assess the episiotomy rate separately for university clinics or for nulliparous and multiparous deliveries. As expected, there seems to be a clear correlation between the instrumental vaginal delivery rate and use of episiotomy (CC 0.47). Still, the very wide range in the use of episiotomy between the hospital size categories (16% to 25%) indicates that its use is not based on uniform protocols. Moreover, the differences are not solely explained by the use of vacuum extraction or by parity or centralization. Considering that episiotomy in non-instrumental vaginal deliveries is primarily performed by midwives, this finding emphasizes that the variance in obstetric practices is not limited to care given or managed by physicians.

Like for episiotomy, also the differences in the use of epidural could be partially explained by differences in the instrumental vaginal delivery rates. However, instrumental vaginal delivery did not explain the variation in episiotomy rates (CC 0.33). We suggest the more active use of epidural in the

larger than in the smaller units is partially explained by better accessibility (anesthesiological responsiveness) and partially by true differences in approaches to pain relief.

It is noteworthy, that neither the episiotomy rate nor the epidural rate were clearly correlated with the obstetric trauma rates within the hospital size categories used (CC 0.17 and -0.24, respectively). It is now established that a mediolateral episiotomy does not increase the risk for obstetric trauma, but there is no consensus about its impact on risk reduction (233-236).

8.3 NEONATAL OUTCOME INDICATORS

The results for all of the studied neonatal outcomes are presented in Table 10 (p. 75) and discussed in detail in the following sections.

8.3.1 Risk for early neonatal mortality is lower in the mid-sized units than in smaller units

The perinatal mortality in the low-risk study population (singleton, non-university births with GA beyond 37 weeks) was 1.39 deaths per 1,000 births during our study period. There was rather a wide inter-unit variation in the perinatal mortality rate from 0 to 3.3 (mean 1.52, SD 0.72). About three-quarters (n=192) of the perinatal deaths (n=259) were stillbirths and one-quarter early neonatal deaths (early neonatal mortality, ENM, n=67). (Table 10).

The perinatal mortality rates – and its components, the early neonatal mortality and stillbirth rates – are shown per unit as a scatter plot in Figure 8. As seen for obstetric trauma, also the perinatal mortality rates show great inter-unit variation within hospital size categories, but when studied by these size categories, the risk for early neonatal death was twice as high in small than in mid-sized units. The rate in the large units was not statistically significantly higher (Table 10).

The stillbirth rates or the perinatal mortality rates did not show statistical differences between the studied hospital categories in the study population. However, in our additional analyses, including also multiple and preterm pregnancies, a lower risk for stillbirth was observed in the large units. This finding further supports the centralization of high-risk pregnancies.

Table 10 *Results for the neonatal indicators. Term (GA 37 weeks or more), singleton deliveries, university clinics excluded (total n=180,368)*

Indicator	Size of birth unit (annual births)	n (population)	rate	OR	95% CI
Perinatal mortality (1/1000 live births)	<1000	39,827	1.61	1.06	0.78-1.43
	1000 to 3000	82,246	1.52	1	
n=259	>3000	58,103	1.20	0.79	0.59-1.06
Early neonatal mortality (1/1000 live births)	<1000	39,827	0.63	2.07*	1.19-3.59
	1000 to 3000	82,246	0.3	1	
n=67	>3000	58,103	0.29	0.96	0.52-1.78
Stillbirths (1/1000 live births)	<1000	39,827	0.98	0.81	0.56-1.17
	1000 to 3000	82,246	1.31	1	
n=192	>3000	58,103	0.91	0.75	0.62-1.41
Umbilical cord pH <7.05	<1000	19,367	1.29 %	1.32*	1.14-1.53
(Hospitals reporting less than 60% excluded,	1000 to 3000	64,907	0.98 %	1	
total n=128,059)	>3000	43,785	1.13 %	1.16*	1.03-1.31
Umbilical cord pH <7.10	<1000	19,367	2.82 %	1.09	0.99-1.20
(Hospitals reporting less than 60% excluded,	1000 to 3000	64,907	2.59 %	1	
total n=128.059)	>3000	43,785	3.29 %	1.27*	1.18-1.36
5-min Apgar <4	<1000	39,866	1.59 %	1.07	0.97-1.18
	1000 to 3000	82,346	1.48 %	1	
	>3000	58,156	1.19 %	0.80*	0.73-0.88
5-min Apgar <7	<1000	39,866	0.23 %	0.72*	0.57-0.92
	1000 to 3000	82,346	0.32 %	1	
	>3000	58,156	0.30 %	0.95	0.78-1.14
Erb's paralysis	<1000	39,866	0.18 %	0.80	0.61-1.04
	1000 to 3000	82,346	0.23 %	1	
	>3000	58,156	0.15 %	0.65*	0.51-0.84
Clavicular fracture	<1000	39,866	1.27 %	1.02	0.92-1.13
	1000 to 3000	82,346	1.24 %	1	
	>3000	58,156	0.57 %	0.46*	0.41-0.52
Birth trauma	<1000 births	39,866	2.17 %	1.13*	1.04-1.22
	1000 to 3000	82,346	1.93 %	1	
	>3000 births	58,156	1.46 %	0.76*	0.70-0.82
Respirator treatment	<1000	39,866	0.25 %	0.99	0.78-1.25
	1000 to 3000	82,346	0.26 %	1	
	>3000	58,156	0.55 %	2.13*	1.79-2.54
Prolonged hospitalization of neonate	<1000	39,866	2.32 %	0.84*	0.78-0.90
	1000 to 3000	82,346	2.78 %	1	
	>3000	58,156	2.96 %	1.06*	1.00-1.13
Proportion of postterm deliveries	<1000	39,866	4.60 %	0.94*	0.89-1.00
	1000 to 3000	82,346	4.88 %	1	
	>3000	58,156	6.62 %	1.36*	1.30-1.42

*) statistically significant with 95% confidence interval

In agreement with our result, in Germany, the smallest units (<500 annual births) had a threefold birthweight-specific mortality risk (17). A study from Norway showed that 2,000 to 3,000 annual births per unit are needed to reduce the risk of neonatal deaths in low-risk deliveries (18). Not supporting the higher risk in smaller units but highlighting the impact ofcentralization of high-risk pregnancies, in Sweden, the neonatal mortality was found to be lower in small than in mid-sized units, but the difference was no longer significant when adjusted with potential confounders (GA, maternal age and BMI, smoking, parental cohabitation) (237).

It is noteworthy that the CS rate and the perinatal mortality rate showed a low to moderate negative correlation in the non-university clinics (CC -0.44), and a relatively strong negative correlation in the units with a CS rate below 15% (CC -0.70). Our study setup was not designed to evaluate whether a low CS rate is associated with high perinatal mortality, and therefore, strong conclusions should not be drawn. However, our finding supports the perception that the very low (<15%) CS rates may be associated with increased perinatal hazards (143, 221).

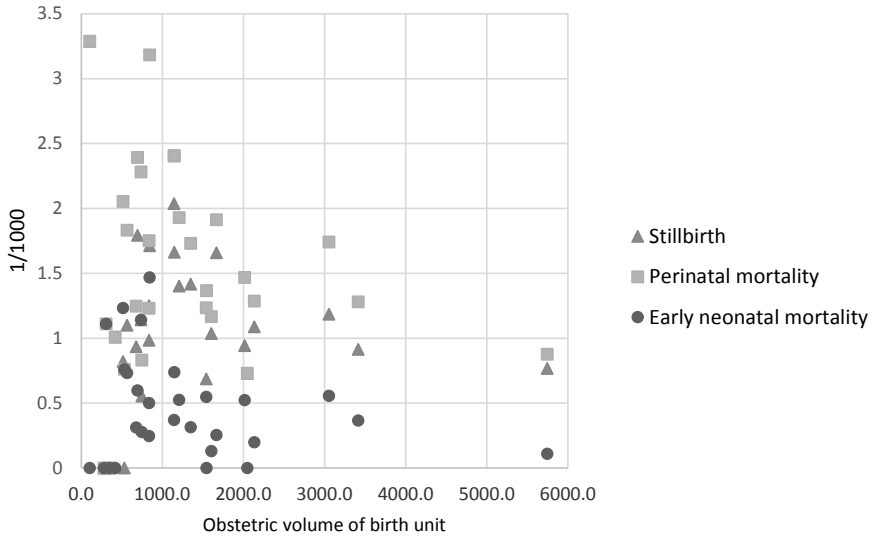


Figure 8 *Perinatal mortality, early neonatal mortality (ENM), and stillbirths in different-sized delivery units.*

8.3.2 Inconsistent results for the two asphyxia indicators: Umbilical cord pH and Apgar score

Our study revealed problems in assessing the frequency of birth asphyxia using the known indicators, 5-min Apgar score and uc-pH; these two outcomes did not show correlation as expected. Again, as fort the other outcomes tested, there was substantial inter-unit variation not clearly explained by the size of the birth unit; polynomial regression curve was the best, yet still poor fit for the model, see figure 9.

The objective measurement, pH, yielded highly consistent rates with the different pH thresholds used (CC >0.90) while the results for Apgar score thresholds 4 and 7 were not as strongly correlated (CC 0.34) in the different birth units. Apgar score is a subjective measurement, and hence, it can be influenced by differences in practices and culture of a birth unit.

The problem with using pH was the very inconsistent reporting between the birth units; uc-pH was reported for 0 to 97% of the births (mean 64%, SD 35%). Low reporting causes a strong risk for bias due to patient selection – in our study, there was a moderate direct correlation between the reporting levels and the rate of adverse outcome (CC 0.60 for pH <7.00, CC 0.70 for pH <7.05 and CC 0.76 for pH<7.10). Hence, the units with high reporting levels are more likely to appear having a high adverse outcome rate. Only the units reporting more than 60% (n=17) were included in the analysis to reduce the confounding effect of varying reporting levels. The results are likely the most reliable for the lowest pH threshold (7.00), since after the units reporting less than 60% were excluded, the reporting levels and outcome rates did not correlate in this group (CC 0.06).

The risk for very low Apgar scores (5-min Apgar <4) was 20% lower in the large than in the mid-sized units. The fact that the risk for very low umbilical cord pH (<7.00) was not increased in these units could be regarded as supporting the Apgar score findings. In contrast, the risk for pH <7.05 was increased in the largest units, but as discussed above, this finding might be confounded by the higher reporting levels in the largest units (Table 10).

The 40% increased risk for very low pH <7.00 in the small units is likely an indicator for an increased risk for birth asphyxia in these units compared to mid-sized units. This interpretation is supported by the significantly increased ENM in the small units in our study as well as by previous studies showing an inverse association between the number of annual deliveries and birth asphyxia (238, 239).

We interpret the observed inconsistency between uc-pH and Apgar in our study to arise mostly from the subjective nature of Apgar scores as an indicator and partly from the inconsistent reporting levels of uc-pH (240, 241).

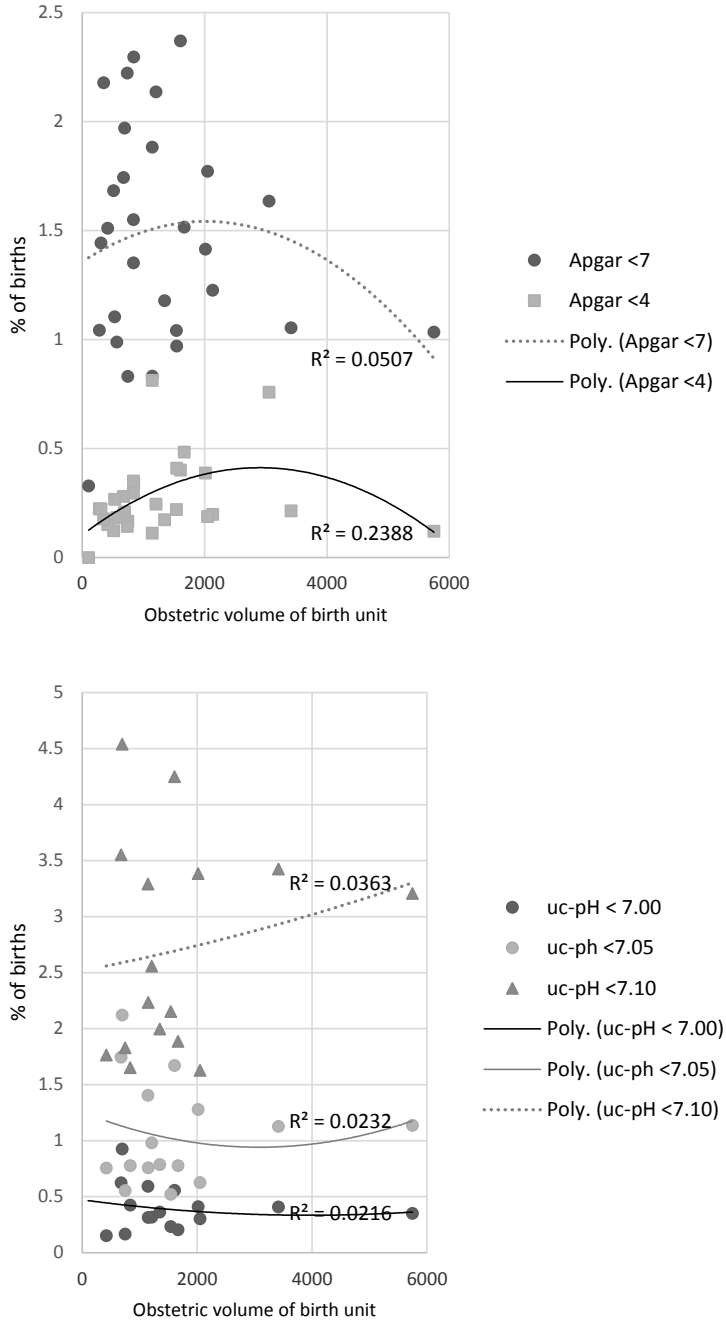


Figure 9 *Neonatal asphyxia indicators in different-sized delivery units. Polynomial trendlines show poor fit for the model.*

8.3.3 Significant differences in neonatal birth trauma rates

Erb's paralysis and clavicular fracture were significantly less likely to occur in large units than in mid-sized units. When using the modified indicator of birth trauma, the incidence was lower in the large and higher in the small units compared with the mid-sized units. (Table 10)

As an indicator, birth trauma includes all discharges with ICD-10 codes (P10-15) for birth trauma as well as for intraventricular non-traumatic hemorrhage (P52) – i.e. the indicator includes also clavicular fracture (P13.4) and Erb's paralysis (P14.0). Since these two diagnoses constitute the majority of all birth trauma cases (in our data 54% and 11%, respectively), the birth trauma indicator is highly correlated with the clavicular fracture rate. When the indicator was tested without clavicular fractures, the small units remained at higher risk, but now also the large units had a higher risk (OR 1.26, 95% CI 1.11-1.43 and OR 1.22 95% CI 1.09-1.36). This result highlights that there are likely to be true differences between different-sized birth units also in the incidence of the more infrequent neonatal traumas. However, the diagnostic and reporting issues related to the rare conditions may confound the result.

Macrosomia, a risk factor for shoulder dystocia (242, 243), did not explain the differences between the birth units in our study. The decreased risk for birth trauma (and clavicular fracture and Erb's paralysis analyzed separately) in the large units could be associated with the larger volume of deliveries leading to more experience in the management of shoulder dystocia. A potential confounder is maternal BMI, which is known to affect the risk for shoulder dystocia (243).

Like obstetric trauma on the mother, birth trauma is a potentially preventable harm in childbirth, but birth trauma has not shown to have high construct nor content validity as a quality indicator, unlike obstetric trauma (129, 244). In our study, the results for birth are congruent with most neonatal outcome indicators enhancing its validity.

8.4 NEONATAL PROCESS INDICATORS

8.4.1 The larger the birth unit, the larger the proportion of postterm births

The proportion of postterm deliveries was significantly higher in large units than in mid-sized units (OR (95% CI) 1.36, 1.31-1.42). The difference appeared to be the more significant the longer the gestation was (Figure 10) with a steady increase in the ORs (at 42^{+0} 1.14, 1.05-1.25, at 42^{+1} 1.64, 1.52-1.77, at 42^{+2} 1.74, 1.57-1.93). In the small units the proportion was

significantly lower than in the mid-sized units at 42^{+1} and 42^{+2} gestational weeks (0.89, 0.80-0.98 and 0.86, 0.75-0.99, respectively).

These differences indicate differences in induction and/or pre-labor CS policies – a more active treatment policy (induction) in the small units and a more conservative treatment policy (expectant management) in the large units. Differences in background factors, parity, and BMI may explain part of the differences in the proportion of postterm births, as discussed in the Review of the literature. However, despite the confounding factors, the results reflect true differences in induction policies and practices highlighting the need for national guidelines on management of prolonged pregnancy.

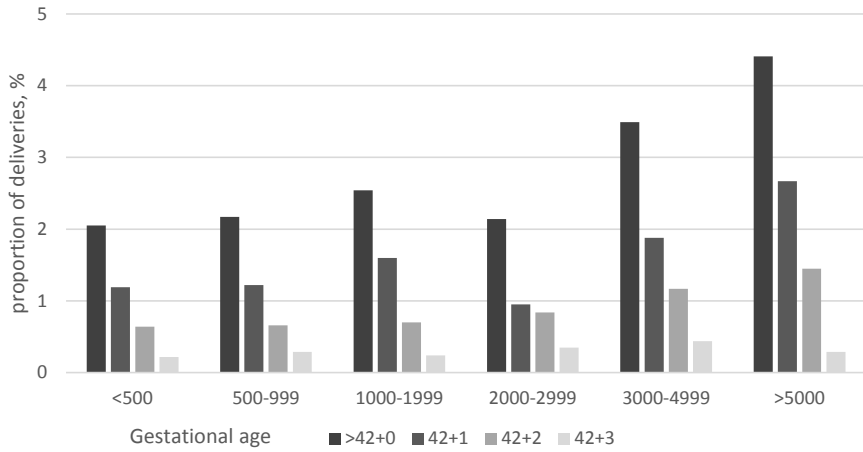


Figure 10 *Proportion of postterm deliveries according to gestational age in each hospital size category (university clinics excluded), %.*

8.4.2 Respirator use on neonate: Does accessibility to treatment improve the outcomes or increase the risk for overtreatment?

The use of a respirator varied significantly according to hospital size, the use being higher the larger the unit. The use was more than twofold in the large units compared with the mid-sized units (OR (95% CI) 2.13, 1.79-2.54) even after excluding the postterm deliveries (2.05, 1.71-2.46). The use of a respirator did not correlate with ENM (outcome indicator for treatment i.e. downstream from the process indicator) or with asphyxia indicators (indication for the treatment, upstream from the process indicator). It did, however, correlate strongly with prolonged hospitalization (CC 0.94).

The results for the respirator treatment aroused concern about whether accessibility to care increases the risk for its overuse. It is known, that in units with more than 3,000 annual deliveries, intensive neonatal care is more readily available than in the mid-sized units – does this create a more active treatment culture without a significant positive impact on the outcomes (21)?

8.4.3 Prolonged hospitalization in large units explained by the large proportion of postterm births

The proportion of newborns still hospitalized after seven days from delivery was significantly lower in the small and higher in the large unit than in the mid-sized units (OR (95% CI) 0.94, 0.89-1.00 and 1.36, 1.30-1.42). For the large units, the finding was explained by the postterm deliveries; not only were there significantly more postterm deliveries, but the postterm newborns were also more often hospitalized for over seven days in the large units (1.50, 1.19-1.89).

8.5 STRENGTHS AND LIMITATIONS OF THE STUDY

Both the strengths and the limitations of the study are related to the registry-based setup. Access to data of the total national population was crucial for the scope of the study, and it highlighted variance in obstetric processes and outcomes. However, despite the high quality of the data, there are several confounding factors warranting cautious interpretation of the results.

The Robson classification enabled reporting for stratified obstetric populations, and hence more accurate comparisons between delivery units. The classification process is not based on information about maternal morbidity and this is a relevant confounder in our study. The centralization of high-risk pregnancies forms a bias for which we were unable to adjust, but for neonatal outcomes, excluding the university clinics limits the confounding effect.

Our study revealed that coarse process measures like CS or instrumental vaginal delivery rate, use of episiotomy and spinal and/or epidural analgesia are size-dependent, and mostly show a bell- or u-shaped association with the size of the delivery unit. Even with a conservative interpretation, this suggests that the volume of a birth unit affects the obstetric processes and the treatment culture. Previously, it has been estimated that in inter-unit comparison, only one-third of the differences in CS rates would be explained by case-mix (245).

Considering the risk factors for obstetric trauma, we regard the case-mix bias to be moderate at the most: many morbidities causing centralization, e.g. pre-eclampsia, intrauterine growth restriction or preterm birth, are not associated with increased odds for obstetric trauma. Maternal BMI may have an impact on obstetric trauma rate (246, 247) and must be regarded as a possible confounder. Gestational diabetes has not been associated with an increased risk for obstetric trauma, except through high birth weight, which has been assessed separately. However, gestational diabetes may increase the risk for birth trauma of the neonate and is a potential confounder for this outcome (248).

We aimed at limiting the analysis of neonatal outcomes to relatively low-risk pregnancies by including births after 37 weeks only and by excluding births at university clinics. Maternal smoking, BMI, parental cohabitation, and pregnancy-related morbidities (e.g. pre-eclampsia) were not controlled for and may have a significant impact on our results. However, as in the Swedish study, most of these factors are related to the increased risk in the largest units, not to the increased risk in the small units seen in our study (237).

Crude analysis, where case-mix is acknowledged but not controlled for, can be justified in some registry setups; these should be regarded as screenings which potentially highlight areas in need of further research. A higher risk of adverse outcome is an important in any case; it either implicates true differences in the delivery of care or differences in the background population. The importance of the latter finding is often neglected, but when an increased risk is explained by case-mix, the finding signals other types of quality issues: Are the high-risk pregnancies centralized optimally? Are there substantial differences in preventive and prenatal care between the catchment areas of the birth units studied? Providing answers to these questions contributes to developing the quality of obstetric care as a continuum between primary and specialized care.

9 DETERMINING BEST PRACTICE: LABOR INDUCTION IN PROLONGED PREGNANCY (IV)

9.1 EFFECT OF LABOR INDUCTION ON MATERNAL AND NEONATAL OUTCOMES

9.1.1 Increased risk for emergency Cesarean section at 41 gestational weeks

According to our data, labor induction increases the risk for emergency CS by nearly 20% around 41 gestational weeks (Table 11). Interestingly, before this, between 40^{+0} and 40^{+6}, labor induction was not associated with an increased risk for CS. Similarly, when approaching 42 gestational weeks (GA 41^{+5}-42^{+0}), labor induction and expectant management did not differ in their risk for emergency CS. The GA trends of the risk for emergency CS in labor induction versus expectant management are presented as a Forest plot in Figure 11

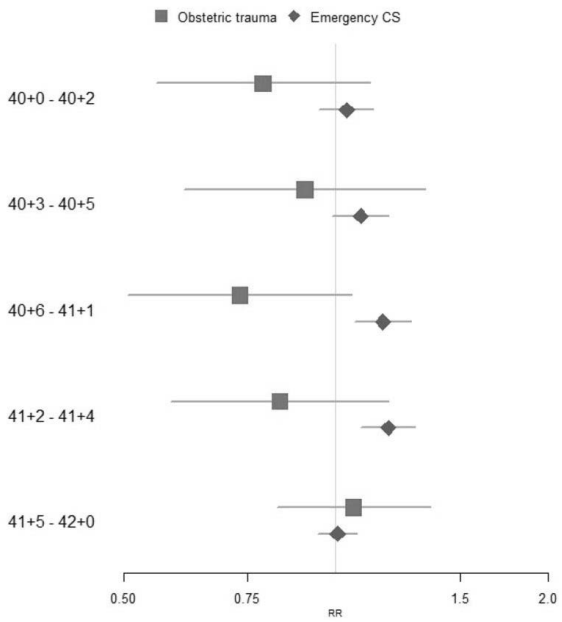


Figure 11 *Forest of maternal outcomes in labor induction versus expectant management. Relative risk and 95% confidence interval anlyzed with Propensity score matched Poisson regression.*

Investigation of the changes in the CS rate among the induced population, in their PS matched cohort, and in ongoing pregnancies for each GA period offers additional information on how the CS risk increases according to GA (Figure 12). The CS rate in ongoing pregnancies increases steadily, but remains substantially lower than in the PS matched cohort in the early GA periods, strengthening the validity of the PS matching.

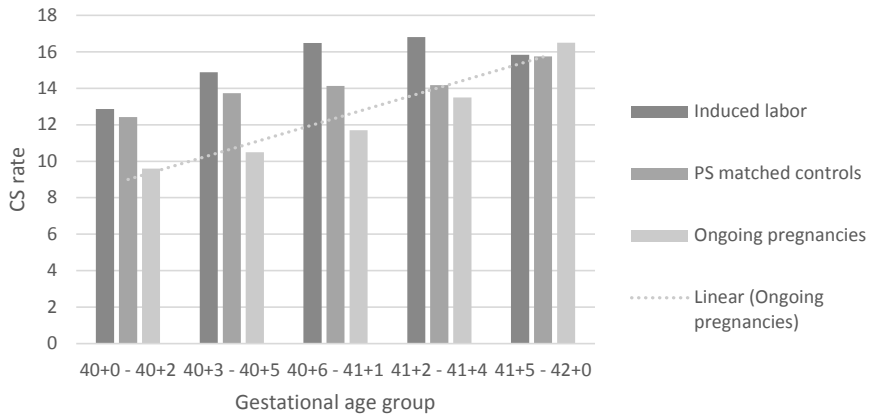


Figure 12 *Cesarean section (CS) rate in induced deliveries (%), in Propensity score (PS) matched controls (expectant management) and in all ongoing pregnancies (linear trendline for CS rate in ongoing pregnancies).*

Based on our results, labor induction seems to increase the risk for any type of operative delivery less than it increases the risk for emergency CS (versus all vaginal deliveries). This indicates that labor induction does not significantly increase the risk for instrumental vaginal delivery or might even decrease it, as has been reported before (177, 249).

Our results on the risk of emergency CS and on any type of operative delivery were not in favor of labor induction at any stage. Previously, most RCTs have reported labor induction to decrease the risk for CS, while cohort and case-control studies show the opposite results: a risk increase by labor induction – and mostly with significantly higher risk estimates than ours (172, 250-253). However, most of the previous cohort studies suffer from methodological issues related to reference group, and to confounding by indication as discussed in the Review of the literature (254). In the retrospective studies of Stock et al. (177) and Danilack et al. (254), the authors focus on the methodology and show how the results differ according to how the study is performed. Their results are in line with ours; induction of labor may increase the risk for emergency CS around gestational week 41.

Table 9 *Relative risk (RR) analyzed with Propensity score (PS) matched Poisson regression in each of the studied gestational periods.*

Gestational age group		*Labor induction*	*Expectant management*		*RR*	*95% CI*
			PS matched controls	All ongoing pregnancies		
		n (%)	*n (%)*	*n (%)*		
40^{+0} - 40^{+2}	*Births/pregnancies*	*6,878*	*6,880*	*205,834*		
	Emergency Cesarean section	885 (12.9)	855 (12.4)	19,794 (9.6)	1.04	0.95-1.13
	Operative delivery†	1,629 (23.7)	1,599 (23.2)	41,156 (20.0)	1.02	0.96-1.08
	Obstetric trauma	56 (0.8)	71 (1.0)	2176 (1.1)	0.79	0.56-1.12
	Perinatal mortality	7 (1.0/1000)	3 (0.4/1000)	81 (0.4/1000)	2.33	0.60-9.02
	Meconium aspiration syndrome	8 (0.1)	20 (0.3)	564 (0.3)	0.40*	0.18-0.91
	5-min Apgar <7	113 (1.6)	119 (1.7)	3,082 (1.5)	0.95	0.74-1.23
	Respirator use on neonate	32 (0.5)	34 (0.5)	917 (0.4)	0.94	0.58-1.52
	Prolonged hospitalization of neonate	305 (4.4)	234 (3.4)	6,046 (2.9)	1.30*	1.10-1.54
40^{+3} - 40^{+5}	*Births/pregnancies*	5,539	5,542	155,809		
	Emergency Cesarean section	825 (14.9)	761 (13.7)	16,324 (10.5)	1.08	0.99-1.19
	Operative delivery†	1,438 (25.9)	1,407 (25.4)	33,250 (21.3)	1.02	0.96-1.09
	Obstetric trauma	47 (0.8)	52 (0.9)	1721 (1.1)	0.90	0.61-1.34
	Perinatal mortality	6 (1.1/1000)	2 (0.4/1000)	63 (0.4/1000)	3.00	0.61-14.86
	Meconium aspiration syndrome	10 (0.2)	23 (0.4)	470 (0.3)	0.44*	0.21-0.91
	5-min Apgar <7	97 (1.7)	89 (1.6)	2,508 (1.6)	1.09	0.82-1.45
	Respirator use on neonate	22 (0.4)	27 (0.5)	22 (0.0)	0.82	0.47-1.43
	Prolonged hospitalization of neonate	254 (4.6)	207 (3.7)	4,699 (3.0)	1.23*	1.03-1.47
40^{+6} - 41^{+1}	*Births/pregnancies*	5,115	5,115	107,160		
	Emergency Cesarean section	843 (16.5)	723 (14.1)	12,572 (11.7)	1.17*	1.06-1.28
	Operative delivery†	1,441 (28.2)	1,364 (26.7)	24,822 (23.2)	1.06	0.99-1.13
	Obstetric trauma	49 (1.0)	67 (1.3)	1243 (1.2)	0.73	0.51-1.06
	Perinatal mortality	1 (0.2/1000)	1 (0.2/1000)	49 (0.5/1000)	1.00	0.06-15.98
	Meconium aspiration syndrome	11 (0.2)	28 (0.5)	376 (0.4)	0.39*	0.20-0.79
	5-min Apgar <7	100 (2.0)	92 (1.8)	1,892 (1.8)	1.09	0.82-1.44
	Respirator use on neonate	23 (0.4)	33 (0.6)	23 (0.0)	0.70	0.41-1.19
	Prolonged hospitalization of neonate	219 (4.3)	188 (3.7)	3,403 (3.2)	1.17	0.96-1.41
41^{+2} - 41^{+4}	*Births / pregnancies*	5,580	5,581	64,629		
	Emergency Cesarean section	938 (16.8)	791 (14.2)	8,735 (13.5)	1.19*	1.09-1.29
	Operative delivery†	1,642 (29.4)	1,449 (26.0)	16,573 (25.6)	1.13*	1.07-1.20
	Obstetric trauma	55 (1.0)	66 (1.2)	797 (1.2)	0.83	0.58-1.19
	Perinatal mortality	2 (0.4/1000)	1 (0.2/1000)	35 (0.5/1000)	2.00	0.18-22.05
	Meconium aspiration syndrome	13 (0.2)	28 (0.5)	273 (0.4)	0.46*	0.24-0.90
	5-min Apgar 5-min Apgar <7	125 (2.2)	103 (1.8)	1251 (2.0)	1.21	0.94-1.57
	Respirator use on neonate	32 (0.6)	31 (0.6)	407 (0.6)	1.03	0.63-1.69
	Prolonged hospitalization of neonate	202 (3.6)	206 (3.7)	2,233 (3.5)	0.98	0.81-1.19
41^{+5} - 42^{+0}	*Births / pregnancies*	10,167	10,167	27,154		
	Emergency Cesarean section	1,610 (15.8)	1,601 (15.7)	4,477 (16.5)	1.01	0.94-1.07
	Operative delivery†	2,850 (28.0)	2,817 (27.7)	8041 (29.6)	1.01	0.97-1.06
	Obstetric trauma	125 (1.2)	118 (1.2)	365 (1.3)	1.06	0.83-1.36
	Perinatal mortality	10 (1.0/1000)	4 (0.4/1000)	10 (0.4/1000)	2.50	0.78-7.97
	Meconium aspiration syndrome	40 (0.4)	43 (0.4)	119 (0.4)	0.93	0.61-1.43
	5-min Apgar <7	218 (2.1)	213 (2.1)	567 (2.1)	1.02	0.85-1.23
	Respirator use on neonate	67 (0.7)	69 (0.7)	188 (0.7)	0.97	0.70-1.36
	Prolonged hospitalization of neonate	412 (4.1)	401 (3.9)	1,028 (3.8)	1.03	0.90-1.18

†) including both emergency Cesarean section and instrumental vaginal delivery,
*) statistically significant with 95% confidence interval

9.1.2 No effect on the risk for obstetric trauma

The risk for obstetric trauma was not affected by labor induction in any GA group. This is in line with our results on obstetric trauma, evaluated using the Robson classification (II); the risk for obstetric trauma in the Robson groups with induced labor (R2a and R4a) was not different from the risk in the respective groups with spontaneous onset of labor (R1 and R3).

Judging by the obstetric trauma rates in different GA periods and the trends in the Forest plot seen in Study IV, GA increases the risk for obstetric trauma, not labor induction. The rates in the ongoing pregnancy cohort increase steadily with increasing GA, similar to the trends in the CS rates with increasing GA. It seems that by labor induction, this risk increase could be avoided, but our results are not significant – perhaps due to the low overall incidence of obstetric trauma in Finland. Previously, Stock et al. showed a statistically significant decrease in the risk for obstetric trauma in labor induction in weeks 39 and 40, but not beyond this (177).

9.1.3 No effect on perinatal mortality

Our results do not support previous findings of substantially decreased perinatal mortality (OR 0.30) with labor induction (28, 168, 177). Our results may be confounded in favor of expectant management due to the exclusion of antepartum stillbirths, which we were not able to assess with the statistical method used. In our study, labor induction did not have a statistically significant effect on stillbirths during labor. The total number of perinatal deaths was low in our study population (n=88) which may have contributed to the statistically insignificant results.

A major concern regarding historical RCTs on labor induction and perinatal mortality is whether the results are applicable in today's obstetric setting. The systematic reviews based on the RCTs do reach statistical significance (28, 168), but considering the original studies mostly date back to the 1990s and 1980s or even earlier, the results might be outdated. A cohort study based on a substantial population, 1.3 million births in Scotland in 1981-2007 (177), reported a similar decrease in perinatal mortality as the reviews. This could be interpreted to increase the validity of the combined data from RCTs. However, considering that pregnancy dating and antenatal fetal surveillance have improved even over the past decade or so, and also the overall perinatal mortality has decreased (255, 256), the effect of labor induction in prolonged pregnancy might not be as favorable as before.

9.1.4 Effect on other neonatal outcomes; decreased risk for meconium aspiration syndrome

Labor induction before 41^{+2} gestational weeks decreases the risk for meconium aspiration syndrome by nearly 60% (Table 11). In contrast, labor induction before 40^{+5} gestational weeks increases the risk for prolonged hospitalization of a neonate by 20-30% (Table11). The risk for low Apgar score was similar in induced deliveries and in their expectantly managed controls in all GA groups studied.

In Denmark, a new, more active induction policy was implemented in 2009 resulting in a steeply decreasing proportion of postterm deliveries (from 5.7% to 2.7% between 2009-10 and 2011-12) (187). In a historical cohort study, the impact of the proactive approach was analyzed by comparing the outcomes on two or three-year periods between 2003 and 2012 with the outcomes in 2000-02 (187). The authors' conclusion is that the new policy has improved the outcomes. However, according to the results presented, most of the improvement took place before 2009-2010 with no further improvement between the last two time-periods studied. Yet, it seems that the policy was only effectively implemented between the last two time-periods (proportion postterm births halved and the term induction rate increased from around 17% to 27%). This change did not improve the results, but instead, seemed to lead to increased NICU admissions.

We regard the conclusions of the Danish study to not to be fully supported by the data presented in their paper, and instead consider their data to support our finding on increased risk for prolonged hospitalization associated with labor induction. It is possible that induction independently increases the neonatal risks resulting in a need for more active or longer postnatal treatment. However, it seems clear that labor induction significantly decreases the risk for meconium aspiration syndrome (257).

9.2 OPTIMAL TIME OF LABOR INDUCTION

The results on the effect of labor induction on maternal and neonatal outcomes were somewhat unexpected: Apart from decreased risk for meconium aspiration syndrome, the study did not show any clear benefits for labor induction in prolonged pregnancy. In light of prior studies, neonatal mortality and morbidity increase significantly in postterm pregnancy (258, 259), but in our data we were not able to see clear neonatal benefit in labor induction after 41 gestational weeks – and where we could see a clear neonatal benefit (on the gestational week 40) there was an increased risk for operative delivery and naturally, a substantial increase in the number of inductions.

Only recently, preliminary results of a large RCT on labor induction have been published; the researches did not find any differences between labor

induction at 41 weeks compared with expectant management until 42 weeks (257). This is an important finding; why intervene if there is no gained benefit? Similar study is ongoing in Sweden, and to be soon launched in Finland, too. These studies should provide us with accurate data to scrutinize labor induction policies (260).

The overall induction rates have increased in all Nordic countries (29). This trend is accompanied by a more proactive approach to prolonged pregnancies with induction often considered preferable to expectant management and the risks involved with postterm pregnancy. This inevitably further increases the intervention rates, and even though the sole intervention would be an induction, the trend contributes to the medicalization of birth. Our study does not support a proactive approach to prolonged pregnancy until close to 42 weeks.

9.3 STRENGTHS AND LIMITATIONS OF THE STUDY

In addition to the substantial population size and high-quality data, the use of the PS method is one of the main assets of the study; its use strengthens the results compared with traditional cohort studies. With the balanced PS matched controls from the "expectantly managed" cohort, we were able to form a study setup where the outcomes of labor induction were compared with the outcomes of cases resembling the result of choosing not to induce. Also, the method permitted us to use a great array of variables as confounding factors. This strongly reduces the bias of the background variables compared with a traditional cohort study setup.

With the three-day GA periods, the study had a clinically relevant approach with regard to best practice protocols; when aiming to adjust the induction policies on a population level, a study assessing the outcomes on each gestational week does not provide precise enough information. Also, the use of three-day study periods revealed a bell-shaped CS risk trend in labor inductions after 40^{+0}, which is an interesting new finding compared to previous studies. Unfortunately, due to the lack of a suitable control group, we were not able to test expectant management beyond 42 gestational weeks. This could have further enhanced the finding on a decreasing CS risk when approaching the postterm period.

Despite the advantages of the PS method, the study also has several limitations. The ratio of induced deliveries to expectantly managed pregnancies naturally varied between the GA groups tested, which could have affected the PS matching, and therefore, might have produced variation in the GA group-specific RRs. Pre-labor CS cases were excluded from the study population (see Section 3.3), and this may confound the results in favor of expectant management.

The study setup did not permit us to use the cervical status (Bishop score) as a background variable. A low score indicates increased risk for labor induction to fail and the use of the score would enable more targeted analysis (261, 262).

Considering the strong evidence for an exponential increase in the neonatal risks with increasing gestational age (28, 168, 177), it is possible the results on weak neonatal benefits in labor inductions near postterm (42 weeks) are biased by some structural elements of the study; e.g. the restricted number of outcomes or the three-day GA period used.

CONCLUSIONS

1 NORDIC CESAREAN SECTION RATES ARE STABILIZING

The use of the Robson classification provided an accurate and focused, yet feasible and explicit way to compare obstetric practices in the Nordic countries with respect to the use of CS. It revealed some clear differences between the Nordic countries, e.g. the differing trends in CS rates among women with a previous Cesarean, as well as similarities, e.g. the increased use of labor induction on nulliparous women.

2 SIZE OF A BIRTH UNIT AFFECTS THE QUALITY OF OBSTETRIC CARE – EVIDENCE FOR CENTRALIZING BIRTHS

Data on outcomes in different-sized delivery units contributes to knowledge management and support health policy decisions. In our study, the number of deliveries correlated with the outcomes, supporting birth centralization, especially from a neonatal perspective.

3 QUALITY OF OBSTETRIC AND PERINATAL CARE SHOULD BE EVALUATED USING A SET OF DIVERSE INDICATORS

Maternal and neonatal outcomes were not aligned in the hospital size categories tested. This highlights the need to measure quality from different perspectives instead of emphasizing the results of a single indicator. Obstetric trauma proved sensitive and feasible in inter-unit comparison. The validity of many neonatal outcome and process measures was weakened by the lack of homogeneous definitions of indicators (e.g. birth trauma, use of intensive care measures on neonate) and varying reporting rates (e.g. for uc-pH).

4 BEST PRACTICE IN PROLONGED PREGNANCY: EXPECTANT MANAGEMENT UNTIL CLOSE TO 42 WEEKS

In this study, the benefits of labor induction in prolonged pregnancy were debatable; the results did not support a proactive approach to prolonged pregnancy until close to 42 weeks. Best practice guidelines suggesting more proactive management should be scrutinized, especially acknowledging the recent trends in induction and CS rates.

FUTURE ASPECTS

It is paramount to design and implement a national perinatal quality program in Finland. The results presented in this thesis, especially the wide range in outcome rates and implications of varying reporting rates, emphasize the need to improve open reporting and benchmarking. Optimally, the indicators would readily direct obstetric processes, e.g. monitoring labor induction rates on nulliparous women could have a positive impact on the induction protocols.

Through internal as well as inter-unit benchmarking of processes and outcomes we may be able to solve some clinical problems difficult to assess through a traditional research setup. The Robson classification has unexploited potential, and the classification should be used for combining the results of CS rate analysis with other outcomes. Such findings as the varying CS rates among women with a previous Cesarean, open possibilities for other meaningful studies on differences in maternal and neonatal outcomes to determine the best practice.

A national perinatal quality program should be relatively easy to implement in a system such as ours, with a long history of high-quality national MBR and ample possibilities for data linkage. Most urgently needed is a balanced and mutually agreed set of indicators to be openly reported for different birth units, levels of care and healthcare regions. Denmark and Sweden serve as examples, having similar healthcare systems, but with established monitoring programs for obstetric care.

In the future, the focus should be put on evaluating indicators for aspects of quality beyond the scope of this thesis, including efficiency (cost-benefit), and furthermore accessibility and patient-centeredness. When large-scale changes are planned for the entire healthcare setting, the importance of open, timely, and valid reports on quality of care increases.

ACKNOWLEDGMENTS

My sincere gratitude goes to

My supervisor Professor *Lasse Lehtonen* for providing a platform for this thesis as part of his research project in the field of patient safety. His experience and expertise in quality of care and patient safety substantially contributed to this thesis.

My supervisor Docent *Anna-Maija Tapper* for providing important clinical insight and for gently yet single-mindedly nudging me forward. She motivated me and believed in me, especially when I needed it most. For all that, she owes my most sincere appreciation and respect.

Docent *Kaarin Mäkikallio* and Docent *Jukka Uotila* for the immense effort they put into the reviewing process leading to a considerably improved end-result.

My author-editor *Carol Ann Pelli* for reviewing the English language of this thesis.

Professor *Mika Gissler*, National Institute for Health and Welfare, without whom this project would not have been possible. He responded in a heartbeat to all of my inquiries on MBR data, never lacking the appropriate figures or a willingness to assist.

Co-authors in Finland for their valuable contribution and support: *Maija Jakobsson* for her energetic and positive presence as well as her ever-so-quick yet poignant remarks; *Jari Petäjä*, for astute and essential insights into the neonatal world; and *Jari Haukka*, for the advanced statistical analyses.

All clinicians and epidemiologists participating in the Nordic Robson research collaboration. Professor *Finn Egil Skjeldestad* for leading the project, and *Ellen Løkkegaard*, *Kari Klungsøyr*, *Alexander Smárason, Thomas Bergholt, Steen Rasmussen*, *Karin Källen, Susanne Albrechtsen*, *Ragnheidur Inga Bjarnadottir*, and *Birna Másdóttir* for important contributions.

Many people not directly related to this project have contributed to it greatly, and I wish to extend my gratitude to all of them. First, thanks to my husband, we relocated to Copenhagen, and leaving clinical work in Finland encouraged me to start the research project, most of which was later carried out during Danish maternity leaves. There were numerous times when the

research kept me busy and preoccupied, and friends and family members took loving care of our *Alma* and *Amos*. I want to express my sincere gratitude for that.

My dear colleagues in Hyvinkää Hospital are thanked for warmly welcoming me back and for providing such a positive and inspirational working environment. The clinic is full of skilled and big-hearted clinicians, all great reminders of why it is a privilege to work in obstetrics and gynecology enduring the long hours and sometimes heart-wrenching cases. This gratitude also extends to all my colleagues in the Women's clinic.

My friends are thanked for keeping me sane and grounded. It is an honor to know so many adventurous, talented and loving ladies who offer vital peer support as well as topics completely outside the fields of medicine and research. Over the years, backpacking together for weeks has evolved into occasional quick chats over (spilled) cups of coffee, but these chats and friendships have become all the more important and meaningful. I value all of you immensely.

I am grateful to my family for unconditional and loving support. I thank my mother *Riitta* for all that she is (and for making sure I am well dressed for the party) and *Antti*, who is no longer here with us, but who taught me to never give up. My father *Tapio* and his wife *Aija* for providing an office, food, and comfort (as well as editing services). My in-laws *Annu* and *Risto* for their help with anything and everything and for being such devoted grandparents. And I thank my siblings, especially *Paula*, who has sympathized during the ups and downs of this journey and put of a lot of thought and effort into making sure it ends with a proper party.

Finally, and with the greatest importance, I want to thank the love of my life, *Lauri-Kustaa*, and our beloved *Alma* and *Amos*, who mean the world to me. *Lauri-Kustaa*, I am forever amazed and continuously inspired by your never-ending optimism, courage and compassion. During this project, you put up with me even when stress made me unbearable. You made – and you keep on making – my everyday better. I love you.

This study was financially supported by
the Doctoral Program in Clinical Research at the University of Helsinki,
the Finnish Cultural Foundation, and the Finnish Society of Perinatology.

Helsinki, August 2017

Aura Pyykönen

REFERENCES

1. Agency for Healthcare Research and Quality (AHRQ). AHRQ Quality Indicators [Report]. 2016 [Available from: https://www.qualityindicators.ahrq.gov/Modules/iqi_resources.aspx.

2. Levine MG, Holroyde J, Woods JR Jr, Siddiqi TA, Scott M, M. M. Birth trauma: incidence and predisposing factors. Obstet Gynecol. 1984;63(6):792-5.

3. Gambone JC, Reiter RC. Elements of a Successful Quality Improvement and Patient Safety Program in Obstetrics and Gynecology. Obstetrics and Gynecology Clinics of North America. 2008;35(1):129-45.

4. Institute of Medicine (IOM). Crossing the quality chasm: A new health system for the 21st century. Washington, D.C: National Academies Press; 2001.

5. World Health Organization (WHO). The World Health Report 2005: Make every mother and child count. Geneva, Switzerland; 2005.

6. Lynn J, Straube BM, Bell KM, Jencks SF, Kambic RT. Using Population Segmentation to Provide Better Health Care for All: The "Bridges to Health" Model. The Milbank Quarterly. 2007;85(2):185-208.

7. Ellsbury DL. Crossing the Quality Chasm in Neonatal-Perinatal Medicine. Clinics in Perinatology. 2010;37(1):1-10.

8. Milbank Memorial Fund. Evidence-based Maternity Care: What It Is and What It Can Achieve. New York, NY; 2008.

9. McKinlay JB. From "Promising Report" to "Standard Procedure": Seven Stages in the Career of a Medical Innovation. The MIlbank Quarterly 1981;59(3):374-411.

10. Forrester King J. A short history of evidence-based obstetric care. Best Practice & Research Clinical Obstetrics & Gynaecology. 2005;19(1):3-14.

11. Cochrane A. Foreword. In: Chalmers I EM, Keirse MJ, editor. Effective Care in Pregnancy and Childbirth. Oxford, New York, Toronto: Oxford University Press; 1987.

12. Carroli G, Mignini L. Episiotomy for vaginal birth. Cochrane Database of Systematic Reviews. 2009(1):73.

13. Alfirevic Z, Devane D, Gyte GML. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. Cochrane Database of Systematic Reviews. 2013(5):139.

14. Crowley P, Chalmers I, Ke1Rse MJNC. The effects of corticosteroid administration before preterm delivery: an overview of the evidence from

controlled trials. BJOG: An International Journal of Obstetrics & Gynaecology. 1990;97(1):11-25.

15. Lasswell S, Barfield W, Rochat R, Blackmon L. Perinatal regionalization for very low-birth-weight and very preterm infants: A meta-analysis. JAMA. 2010;304(9):992-1000.

16. Ryan G. Toward improving the outcome of pregnancy. Recommendations for the regional development of perinatal health services. Obstet Gynecol. 1975;46(4):375-84.

17. Günther Heller DKR, Rainer Schnell, Björn Misselwitz, Wolfgang Künzele and Stephan Schmidt. Are we regionalized enough? Early-neonatal deaths in low-risk births by the size of delivery units in Hesse, Germany 1990–1999. International Journal of Epidemiology. 2002;31:1061-8.

18. Moster D, Lie RT, Markestad T. Relation between size of delivery unit and neonatal death in low risk deliveries: population based study. Archives of Disease in Childhood - Fetal and Neonatal Edition. 1999;80:F221-F5.

19. Hemminki E, Heino A, Gissler M. Should births be centralised in higher level hospitals? Experiences from regionalised health care in Finland. BJOG: An International Journal of Obstetrics & Gynaecology. 2011;118(10):1186-95.

20. Snowden JM, Cheng YW, Emeis CL, Caughey AB. The impact of hospital obstetric volume on maternal outcomes in term, non-low-birthweight pregnancies. Am J Obstet Gynecol. 2015;212(3):380.e1-.e9.

21. Wagner M. Fish can't see water: the need to humanize birth. International Journal of Gynecology & Obstetrics. 2001;75, Supplement 1:S25-S37.

22. OECD. Health at a Glance 2011: OECD Indicators. Report. 2011.

23. Melman S, Schoorel ECN, de Boer K, Burggraaf H, Derks JB, van Dijk D, van Dillen J, Dirksen CD, Duvekot JJ, Franx A, Hasaart THM, Huisjes AJM, Kolkman D, van Kuijk S, Kwee A, Mol BW, van Pampus MG, de Roon-Immerzeel A, van Roosmalen JJM, Roumen FJME, Smid-Koopman E, Smits L, Spaans WA, Visser H, van Wijngaarden WJ, Willekes C, Wouters MGAJ, Nijhuis JG, Hermens RPMG, Scheepers HCJ. Development and Measurement of Guidelines-Based Quality Indicators of Caesarean Section Care in the Netherlands: A RAND-Modified Delphi Procedure and Retrospective Medical Chart Review. PLoS ONE. 2016;11(1):e0145771.

24. Kesmodel US, JøLving LR. Measuring and improving quality in obstetrics – the implementation of national indicators in Denmark. 2011.

25. Euro-Peristat project with Scope and Eurocat. European Perinatal Health Report - Health and Care of Pregnant Women and Babies in Europe 2010. Report. 2013.

26. Betrán AP YJ, Moller AB, Zhang J, Gülmezoglu AM, et al. The Increasing Trend in Caesarean Section Rates: Global, Regional and National Estimates: 1990-2014. PLoS ONE. 2016;11(2):e0148343.

27. Saccone G, Berghella V. Induction of labor at full term in uncomplicated singleton gestations: a systematic review and metaanalysis of randomized controlled trials. Am J Obstet Gynecol. 2015;213(5):629-36.

28. Gülmezoglu AM, Crowther CA, Middleton P, Heatley E. Induction of labour for improving birth outcomes for women at or beyond term. Cochrane Database of Systematic Reviews. 2012;6.

29. National Institute for Health and Welfare (THL). Official Statistics of Finland: Perinatal statistics - parturients, deliveries and newborns 2015. Helsinki, Finland; 2016 2016.

30. Teperi J. Multi method approach to the assessment of data quality in the Finnish Medical Birth Registry. Journal of Epidemiology and Community Health. 1993;47(3):242-7.

31. Gissler M, Teperi J, Hemminki E, Meriläinen J. Data quality after restructuring a national medical registry. Scandinavian Journal of Social Medicine. 1995;23(1):75-80.

32. Campbell SM, Roland MO, Buetow SA. Defining quality of care. Social Science & Medicine. 2000;51(11):1611-25.

33. Institute of Medicine (IOM). Medicare - A strategy for Quality Assurance. Washington D.C.: National Academies Press; 1990.

34. Council of Europe. The development and implementation of quality improvement systems (QIS) in health care: recommendation No. R (97) 17 and explanatory memorandum. . 1998.

35. World Health Organization (WHO). Quality of care: a process for making strategic choices in health systems. Geneva: World Health Organization; 2006.

36. Expert Panel on effective ways of investing in Health. Future EU Agenda on quality of health care witha special emphasis on patient safety. Brussels, Belgium; 2014.

37. Institute of Medicine (IOM). Performance Measurement: Accelerating Improvement. Washington, D.C.: National Academies Press; 2005.

38. Emanuel L, Berwick D, Conway J, Combes J, Hatlie M, Leape L, Reason J, Schyve P, Vincent C, Walton M. What exactly is patient safety? In: Kerm Henriksen, James B Battles, Margaret A Keyes, Mary L Grady,

editors. Advances in Patient Safety: New Directions and Alternative Approaches. Rockville, MD: Agency for Healthcare Research and Quality; 2008.

39. Byers JF, White SV. Patient safety: Principles and practice. New York: Springer Publishing Company; 2004.

40. Institute of Medicine (IOM). Patient Safety: Achieving a New Standard for Care. Washington D.C.: National Academies Press; 2004.

41. Institute of Medicine (IOM). To Err is Human: Building a Safer Health System. Washington D.C.: National Academies Press; 1999.

42. Sheingold BH, Hahn JA. The history of healthcare quality: The first 100 years 1860–1960. International Journal of Africa Nursing Sciences. 2014;1:18-22.

43. Shewhart W. Economic control of quality of manufactured product. New York, NY: Van Nostrand; 1931.

44. Juran JM. Juran on Planning for Quality. New York: Free Press; 1988.

45. Luce JM, Bindman AB, Lee PR. A brief history of health care quality assessment and improvement in the United States. Western Journal of Medicine. 1994;160(3):263-8.

46. Donabedian A. Evaluating the Quality of Medical Care. [Article]. The Milbank Quarterly. 1966;44(3):166-203.

47. Irvine D, Donaldson L. Quality and standards in health care. Proc R Soc Edin. 1993;101B:1-30.

48. Steffen GE. Quality Medical Care: A Definition. JAMA: The Journal of the American Medical Association. 1988;260(1):56-61.

49. Tarlov AR, Ware JE, Jr., Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study: An Application of Methods for Monitoring the Results of Medical Care. JAMA: The Journal of the American Medical Association. 1989;262(7):925-30.

50. Mainz J. Defining and classifying clinical indicators for quality improvement. International Journal for Quality in Health Care. 2003;15(6):523-30.

51. Tripathi V. A literature review of quantitative indicators to measure the quality of labor and delivery care. International Journal of Gynecology and Obstetrics. 2016;132(2):139-45.

52. Cochrane A. Effectiveness and Efficiency. Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust; 1972.

53. Bhatt A. Evolution of Clinical Research: A History Before and Beyond James Lind. Perspectives in Clinical Research. 2010;1(1):6-10.

54. Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. . Washington, DC: National Academy Press; 2001.

55. Porter ME. A Strategy for Health Care Reform — Toward a Value-Based System. New England Journal of Medicine. 2009;361(2):109-12.

56. Main EK, Bloomfield L, Hunt G, for the Sutter Health FPaDCIC. Development of a large-scale obstetric quality-improvement program that focused on the nulliparous patient at term. Am J Obstet Gynecol. 2004;190(6):1747-56.

57. Beattie M, Lauder W, Atherton I, Murphy DJ. Instruments to measure patient experience of health care quality in hospitals: a systematic review protocol. Systematic Reviews. 2014;3:4-.

58. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychological Bulletin. 1955;52(4):281-302.

59. McGlynn EA. Selecting common measures of quality and system performance. Med Care. 2003;41(1 Suppl):I39-47.

60. Brook RH, McGlynn EA, Cleary PD. Measuring Quality of Care. New England Journal of Medicine. 1996;335(13):966-70.

61. Fischer C, Lingsma HF, Anema HA, Kievit J, Steyerberg EW, Klazinga N. Testing the construct validity of hospital care quality indicators: a case study on hip replacement. BMC Health Services Research. 2016;16:551.

62. Janakiraman VMD, Ecker JMD. Quality in Obstetric Care: Measuring What Matters. [Editorial]. Obstet Gynecol. 2010;116(3):728-32.

63. Long A, Fairfield G. Confusion of levels in monitoring outcomes and/or process. The Lancet. 1996;347(9015):1572.

64. Iezzoni LI. Risk Adjustment for Measuring Healthcare Outcomes. Chicago: Ill Health Administration Press; 1994.

65. Keeler EB, Park RE, Bell RM, Gifford DS, Keesey J. Adjusting cesarean delivery rates for case mix. Health Services Research. 1997;32(4):511-28.

66. Iezzoni LI. Assessing Quality Using Administrative Data. Annals of Internal Medicine. 1997;127:666-74.

67. Rubin H, Pronovost P, Diette G. The advantages and disadvantages of process-based measures of health care quality. Int J Qual Health Care. 2001;13(6):469-74.

68. Hummel J, & Huitt, W. What you measure is what you get. GaASCD Newsletter: The Reporter. 1994;10-11.

69. Helmreich RL. On error management: lessons from aviation. BMJ : British Medical Journal. 2000;320(7237):781-5.

70. Berwick DM. A primer on leading the improvement of systems. BMJ. 1996;312(7031):619-22.

71. Reason J. A systems approach to organizational error. Ergonomics. 1995;38(8):1708-21.

72. Ruuhilehto K KM, Keistinen T, Kinnunen M, Vuorenkoski L, Wallenius J. HaiPro - what was learned from patient safety incidents in Finnish health care units in 2007 to 2009? Duodecim. 2011;127(10):1003-40.

73. Knuuttila J, Tamminen A. Helsinki: National Agency for Medicines; 2004.

74. Ministry of Social Affairs and Health. Quality Management in Social and Health Care on the 21st Century. National Recommendation. Helsinki, Finland; 1999.

75. Health Care Act, (2010).

76. Spector JM, Agrawal P, Kodkany B, Lipsitz S, Lashoher A, Dziekan G, Bahl R, Merialdi M, Mathai M, Lemer C, Gawande A. Improving Quality of Care for Maternal and Newborn Health: Prospective Pilot Study of the WHO Safe Childbirth Checklist Program. PLOS ONE. 2012;7(5):e35151.

77. Miller S, Cordero M, Coleman AL, Figueroa J, Brito-Anderson S, Dabagh R, Calderon V, Cáceres F, Fernandez AJ, Nunez M. Quality of care in institutionalized deliveries: The paradox of the Dominican Republic. International Journal of Gynecology and Obstetrics. 2003;82(1):89-103.

78. Main E, Menard M, Kathryn Md MPH. Maternal Mortality: Time for National Action. Obstet Gynecol. 2013;122(4):735-6.

79. Alkema L, Chou D, Hogan D, Zhang S, Moller A-B, Gemmill A, Fat DM, Boerma T, Temmerman M, Mathers C, Say L. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group. The Lancet.387(10017):462-74.

80. WHO UNICEF UNFPA the World Bank and the United Nations Population Division. Trends in Maternal Mortality: 1990 to 2013. Geneva: WHO Press; 2014.

81. Berg C, Md MPH, Callaghan W, Md MPH, Syverson C, Cnm MN, Henderson Z. Pregnancy-Related Mortality in the United States, 1998 to 2005. Obstet Gynecol. 2010;116(6):1302-9.

82. Knight M, Kenyon S, Brocklehurst P, Neilson J, Shakespeare J, (Eds) JJK, UK oboM-. Saving Lives, Improving Mothers' Care. Lessons learned to inform future maternity care from the UK and Ireland. Confidential Enquiries into Maternal Deaths and Morbidity 2009-2012. University of Oxford; 2014.

83. Glass NE, Schulkin J, Chamany S, Riley LE, Schuchat A, Schrag S. Opportunities to reduce overuse of antibiotics for perinatal group B streptococcal disease prevention and management of preterm premature rupture of membranes. Infectious Diseases in Obstetrics and Gynecology. 2005;13(1):5-10.

84. Bingham D, Ruhl C, Cockey CD. Don't Rush Me . . . Go the Full 40: AWHONN's Public Health Campaign Promotes Spontaneous Labor and Normal Birth to Reduce Overuse of Inductions and Cesareans. The Journal of Perinatal Education. 2013;22(4):189-93.

85. Viswanathan M HK, Palmieri R, Lux L, Swinson T, Lohr KN, Gartlehner G, Thorp J Jr. The Use of Episiotomy in Obstetrical Care: A Systematic Review: Summary. AHRQ Evidence Report Summaries. 1998-2005. Rockville, MD: Agency for Healthcare Research and Quality 2005.

86. World Bank. World Bank Open Data 2017 [Available from: http://data.worldbank.org/.

87. Räisänen S, Vehviläinen-Julkunen K, Gissler M, Heinonen S. Up to seven-fold inter-hospital differences in obstetric anal sphincter injury rates- A birth register-based study in Finland. BMC Research Notes. 2010;345(3).

88. Gissler M, Merilainen J, Vuori E, Hemminki E. Register based monitoring shows decreasing socioeconomic differences in Finnish perinatal health. Journal of Epidemiology and Community Health. 2003;57(6):433-9.

89. Talala KH, Tommi; Martelin, Tuija; Karvonen, Sakari; Mäki-Opas, Tomi; Marderbacka, Kristiina; Suvisaari, Jaana; Sainio, Päivi; Rissanen, Harri; Ruokolainen, Otto; Heloma, Antero; Koskinen, Seppo. Time trends in educational disparities in health and well-being in the Finnish population from 2000 to 2011. Suomen Lääkärilehti (Finnish Medical Journal). 2014;69.

90. Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation. British Medical Journal. 1948;2(4582):769-82.

91. Chalmers I EM, Keirse MJNC. Effective Care in Pregnancy and Childbirth. Oxford, New York, Toronto: Oxford University Press; 1989.

92. Johansen KS. The OBSQID Project - Obstetrical Quality Development through Integrated Use of Telematics. In: F.H. Roger France JNvG, K. Staehr Johansen, editor. Studies in Health Technology and Informatics. 141994.

93. Johansen KS, Hod M. Quality development in perinatal care — the OBSQID project. International Journal of Gynecology & Obstetrics. 1999;64(2):167-72.

94. Specifications Manual for Joint Commission National Quality Measures (v2016A1) [Internet]. 2016 [cited Jan 2017]. Available from: https://manual.jointcommission.org/releases/TJC2016A1/PerinatalCare.html.

95. Royal College of Obstetricians and Gynaecologists (RCOG). Patterns of maternity care in English NHS trusts 2013/14. London: Royal College of Obstetricians and Gynaecologists; 2016.

96. United Nations (UN). Millennium Development Goals Indicators: United Nations Statistics Division Published 2008. Accessed 2017; 2008 [Available from: http://mdgs.un.org/unsd/mdg/Host.aspx?Content=Indicators/OfficialList.htm.

97. United Nations. MDG 4: Reduce Child Mortality [Available from: http://www.mdgmonitor.org/mdg-4-reduce-child-mortality/.

98. United Nations. MDG 5: Improve maternal health [Available from: http://www.mdgmonitor.org/mdg-5-improve-maternal-health/.

99. Roser M. Maternal Mortality 2016 [Available from: https://ourworldindata.org/maternal-mortality/.

100. National Institute for Health and Welfare (THL). Quality and Patient Safety Indicators in Obstetrics. A review on how quality and patient safety have been measured in high-income countries with a recommendation on a core set of indicators for the Nordic countries. Helsinki, Finland; 2015. Contract No.: ISBN 978-952-302-545-5.

101. Andersson CB, Flems C, Kesmodel US. The Danish National Quality Database for Births. Clinical Epidemiology. 2016;8:595-9.

102. Socialstyrelsen. Hälso- och sjukvård jämförelser mellan landsting: Del 2. Indikatorer om sjukdomar och behandlingar. Socialstyrelsen; 2014. Report No.: ISBN 978-91-7555-241-5.

103. Reisner DP, Landers S. Collaboration Between Obstetricians and Neonatologists: Perinatal Safety Programs and Improved Clinical Outcomes. Clinics in Perinatology. 2010;37(1):179-88.

104. World Health Organisation (WHO). Having a Baby in Europe. European Regional Office; 1985.

105. Lothian JA. Safe, Healthy Birth: What Every Pregnant Woman Needs to Know. The Journal of Perinatal Education. 2009;18(3):48-54.

106. Strong TH. Expecting Trouble: What Expectant Parents Should Know About Prenatal Care in America2002.

107. Räisänen S, Lehto SM, Nielsen HS, Gissler M, Kramer MR, Heinonen S. Fear of childbirth in nulliparous and multiparous women: a population-based analysis of all singleton births in Finland in 1997–2010.

BJOG: An International Journal of Obstetrics & Gynaecology. 2014;121(8):965-70.

108. Gaskin IM. Birth Matters. A Midwife's Manifesta. Glasgow, Scotland: Thomson Litho; 2011.

109. Wolfe MW. Mother and child: a multi-determinant model for maternal and infant health outcomes in urban, low-income communities and the effectiveness of prenatal care and other interventions. Journal of Public & International Affairs. 2006;17.

110. Bailit JLMDMPH. Measuring the Quality of Inpatient Obstetrical Care. [Review]. Obstetrical & Gynecological Survey March. 2007;62(3):207-13.

111. Mann S, Pratt S, Gluck P, Nielsen P, Risser D, Greenberg P, Marcus R, Goldman M, Shapiro D, Pearlman M, Sachs B. Assessing Quality in Obstetrical Care: Development of Standardized Measures. Journal on Quality and Patient Safety. 2006;32(9):497-505.

112. Institute of Medicine (IOM). Envisioning the National Health Care Quality Report. Washington (DC): National Academies Press (US); 2001 [

113. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A Method for Continuous Quality Improvement in Health. Healthcare Policy. 2012;7(4):e101-e19.

114. Woitha K, Van Beek K, Ahmed N, Jaspers B, Mollard JM, Ahmedzai SH, Hasselaar J, Menten J, Vissers K, Engels Y. Validation of quality indicators for the organization of palliative care: A modified RAND Delphi study in seven European countries (the Europall project). Palliative Medicine. 2013;28(2):121-9.

115. Thakar R EE. Prevention of perineal trauma. In: Sultan AH TR, Fenner DE, editor. Perineal and anal sphincter trauma. 1st ed. London, GB: Springer-Verlag; 2009.

116. Handa VL DB, Gilbert WM. Obstetric anal sphincter lacerations. Obstet Gynecol. 2001;98:225-30.

117. Raleigh VS, Cooper J, Bremner SA, Scobie S. Patient safety indicators for England from hospital administrative data: case-control analysis and comparison with US data. BMJ. 2008;337.

118. Sultan AH, Kamm MA, Hudson CN, Bartram CI. Third degree obstetric anal sphincter tears: risk factors and outcome of primary repair. BMJ. 1994;308:887-91.

119. Pollack J, Nordenstam J, Brismar S, Lopez A, Altman D, Zetterstrom J. Anal incontinence after vaginal delivery: a five-year prospective cohort study. Obstet Gynecol. 2004;104(6):1397-402.

120. Nichols CM, Nam M, Ramakrishnan V, Lamb EH, Currie N. Anal sphincter defects and bowel symptoms in women with and without recognized anal sphincter trauma. Am J Obstet Gynecol. 2006;194(5):1450-4.

121. Laine K, Pirhonen T, Rolland R, Pirhonen J. Decreasing the incidence of anal sphincter tears during delivery. Obstet Gynecol. 2008;111(5):1053-7.

122. Laine K, Skjeldestad FE, Sandvik L, Staff AC. Incidence of obstetric anal sphincter injuries after training to protect the perineum: cohort study. BMJ Open. 2012;2.

123. Naidu M, Sultan AH, Thakar R. Reducing obstetric anal sphincter injuries using perineal support: our preliminary experience. International Urogynecology Journal. 2016:1-9.

124. Downey JR, Hernandez-Boussard T, Banka G, Morton JM. Is Patient Safety Improving? National Trends in Patient Safety Indicators: 1998–2007. Health Services Research. 2012;47(1pt2):414-30.

125. Grobman WA, Feinglass J, Murthy S. Are the Agency for Healthcare Research and Quality obstetric trauma indicators valid measures of hospital safety? Am J Obstet Gynecol. 2006;195(3):868-74.

126. Apgar V. A Proposal for a New Method of Evaluation of the Newborn Infant. Anesthesia & Analgesia. 1953;32(4):260-7.

127. Apgar V, Holaday DA, James L, Weisbrot IM, Berrien C. Evaluation of the newborn infant-second report. Journal of the American Medical Association. 1958;168(15):1985-8.

128. Armstrong L, Stenson BJ. Use of umbilical cord blood gas analysis in the assessment of the newborn. Archives of Disease in Childhood Fetal and Neonatal Edition. 2007;92(6):F430-F4.

129. Miller MR, Pronovost P, Donithan M, et al. Relationship between performance measurement and accreditation: implications for quality of care and patient safety. Am J Med Qual. 2005;20:239-52.

130. National Center for Health Statistics. International Classification of Diseases,Ninth Revision, Clinical Modification (ICD-9-CM) [Available from: International Classification of Diseases,Ninth Revision, Clinical Modification (ICD-9-CM).

131. Horbar JD. The Vermont Oxford Network: Evidence-Based Quality Improvement for Neonatology. Pediatrics. 1999;103(Supplement E1):350.

132. Horbar JD, Edwards EM, Greenberg LT, et al. Variation in performance of neonatal intensive care units in the united states. JAMA Pediatrics. 2017;171(3):e164396.

133. Dukhovny D, Pursley DM, Kirpalani HM, Horbar JH, Zupancic JAF. Evidence, Quality, and Waste: Solving the Value Equation in Neonatology. Pediatrics. 2016;137(3).

134. Guihard P, Blondel B. Trends in risk factors for caesarean sections in France between 1981 and 1995: lessons for reducing the rates in the future. BJOG: An International Journal of Obstetrics & Gynaecology. 2001;108(1):48-55.

135. Queenan JT. How to Stop the Relentless Rise in Cesarean Deliveries. [Editorial]. Obstet Gynecol. 2011;118:1422-3.

136. Betrán AP, Merialdi M, Lauer JA, Bing-Shun W, Thomas J, Van Look P, Wagner M. Rates of caesarean section: Analysis of global, regional and national estimates. Paediatric and Perinatal Epidemiology. 2007;21(2):98-113.

137. Macfarlane AJ, B Blondel B, Mohangoo AD, Cuttini M, Nijhuis J, Novak Z, Olafsdottir HS, Zeitlin J, Committee E-PS. Wide differences in mode of delivery within Europe: risk-stratified analyses of aggregated routine data from the Euro-Peristat study. BJOG: An International Journal of Obstetrics & Gynaecology. 2016;123(4):559–68.

138. World Health Organisation (WHO). Appropriate technology for birth. The Lancet. 1985;2:436-7.

139. World Health Organization (WHO). Statement on Caesarean Section Rates. : WHO Press; 2015 [Available from: http://apps.who.int/iris/bitstream/10665/161442/1/WHO_RHR_15.02_eng.pdf?ua=1.

140. The Joint Commission. America's Hospitals: Improving Quality and Safety. The Joint Commission's Annual Report. 2014.

141. Howell EA, Zeitlin J, Hebert PL, Balbierz A, Egorova N. The Association between Hospital-level Obstetric Quality Indicators and Maternal and Neonatal Morbidity. JAMA. 2014;312(15):1531-41.

142. Robson MS. Classification of caesarean sections. Fetal and Maternal Medicine Review. 2001;12(1):23-39.

143. Betran AP, Torloni MR, Zhang J, Ye J, Mikolajczyk R, Deneux-Tharaux C, Oladapo OT, Souza JP, Tunçalp Ö, Vogel JP, Gülmezoglu AM. What is the optimal rate of caesarean section at population level? A systematic review of ecologic studies. Reproductive Health. 2015;12(1):1-10.

144. Souza JP, Betran AP, Dumont A, de Mucio B, Gibbs Pickens CM, Deneux-Tharaux C, Ortiz-Panozo E, Sullivan E, Ota E, Togoobaatar G, Carroli G, Knight H, Zhang J, Cecatti JG, Vogel JP, Jayaratne K, Leal MC, Gissler M, Morisaki N, Lack N, Oladapo OT, Tunçalp Ö, Lumbiganon P, Mori R, Quintana S, Costa Passos AD, Marcolin AC, Zongo A, Blondel B, Hernández B, Hogue CJ, Prunet C, Landman C, Ochir C, Cuesta C,

Pileggi-Castro C, Walker D, Alves D, Abalos E, Moises ECD, Vieira EM, Duarte G, Perdona G, Gurol-Urganci I, Takahiko K, Moscovici L, Campodonico L, Oliveira-Ciabati L, Laopaiboon M, Danansuriya M, Nakamura-Pereira M, Costa ML, Torloni MR, Kramer MR, Borges P, Olkhanud PB, Pérez-Cuevas R, Agampodi SB, Mittal S, Serruya S, Bataglia V, Li Z, Temmerman M, Gülmezoglu AM. A global reference for caesarean section rates (C-Model): a multicountry cross-sectional study. BJOG : an international journal of obstetrics and gynaecology. 2016;123(3):427-36.

145. Bjarnadottir R, Smarason A. Poster Session Abstracts. Acta Obstetricia et Gynecologica Scandinavica. 2012;91:59--149.

146. Scarella A, Chamy V, Sepúlveda M, Belizán JM. Medical audit using the Ten Group Classification System and its impact on the cesarean section rate. Eur J Obstet Gynecol Reprod Biol. 2011;154(2):136-40.

147. Vogel JP, Betran AP, Vindevoghel N, Souza JP, Torloni MR, Zhang J. Use of the Robson classification to assess caesarean section trends in 21 countries: a secondary analysis of two WHO multicountry surveys. The Lancet Global health. 2015;3.

148. Torloni MR, Betran AP, Souza JP, Widmer M, Allen T, Gulmezoglu M, Merialdi M. Classifications for Cesarean Section: A Systematic Review. PLoS ONE: - Public Library of Science; 2011. p. e14566.

149. Betran AP, Vindevoghel N, Souza JP, Gulmezoglu AM, Torloni MR. A Systematic Review of the Robson Classification for Caesarean Section: What Works, Doesn't Work and How to Improve It. PLoS One. 2014;9.

150. Phibbs CS, Bronstein JM, Buxton E, Phibbs RH. The effects of patient volume and level of care at the hospital of birth on neonatal mortality. JAMA-J Am Med Assoc. 1996;276(13):1054-9.

151. Cifuentes J, Bronstein J, Phibbs CS, Phibbs RH, Schmitt SK, Carlo WA. Mortality in Low Birth Weight Infants According to Level of Neonatal Care at Hospital of Birth. Pediatrics. 2002;109(5):745.

152. Milland M, Christoffersen JK, Hedegaard M. The size of the labor wards: is bigger better when it comes to patient safety? Acta Obstet Gynecol Scand. 2013;92(11):1271-6.

153. World Health Organization (WHO). Standards for improving quality of maternal and newborn care in health facilities. Geneva, Switzerland; 2016.

154. Srinivas SKM, MSCE, Fager CM, Lorch SAM, MSCE. Evaluating Risk-Adjusted Cesarean Delivery Rate as a Measure of Obstetric Quality. Obstetrics & Gynecology. 2010;115(5):1007-13.

155. Lain SJ, Algert CS, Nassar N, Bowen JR, Roberts CL. Incidence of Severe Adverse Neonatal Outcomes: Use of a Composite Indicator in a Population Cohort. Maternal and Child Health Journal. 2012;16(3):600-8.

156. Pettker CM, Thung SF, Norwitz ER, Buhimschi CS, Raab CA, Copel JA, Kuczynski E, Lockwood CJ, Funai EF. Impact of a comprehensive patient safety strategy on obstetric adverse events. American Journal of Obstetrics and Gynecology. 2009;200(5):492.e1-.e8.

157. Emily Hamilton M, Samuel Smith M, Dorothy Berry R, DFASHRM; , Omer Ben-Yoseph M, Dan O'Keeffe M, Eric Knox M. Revisiting the Perinatal Adverse Outcome Index. Patient Safety and Quality Healthcare. 2011(May/June).

158. Papanicolas I, Smith PC, Mossialos E. Methodological paper: What are the methodological issues related to measuring health care performance? In: London School of Economics and Political Science, editor. European Commission2008.

159. Chalmers B, Porter R. Assessing effective care in normal labor: The Bologna score. Birth. 2001;28(2):79-83.

160. Arjakari O, Mallat K. The birth experience with the evaluation of the Visual analogue scale. Helsinki: Metropolia, University of Applied Sciences; 2012.

161. Rouhe H, Salmela-Aro K, Halmesmäki E, Saisto T. Fear of childbirth according to parity, gestational age, and obstetric history. BJOG: An International Journal of Obstetrics & Gynaecology. 2009;116(1):67-73.

162. Committee on Perinatal Health. Toward Improving the Outcome of Pregnancy: Enhancing Perinatal Health Through Quality, Safety and Performance Initiatives. White Plains, NY; 2010.

163. Hogle KL, Kilburn L, Hewson S, Gafni A, Wall R, Hannah ME. Impact of the International Term Breech Trial on Clinical Practice and Concerns: A Survey of Centre Collaborators. Journal of Obstetrics and Gynaecology Canada. 2003;25(1):14-6.

164. Hannah ME, Hannah WJ, Hewson SA, Hodnett ED, Saigal S, Willan AR. Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial. The Lancet. 2000;356(9239):1375-83.

165. Glezerman M. Five years to the term breech trial: The rise and fall of a randomized controlled trial. Am J Obstet Gynecol. 2006;194(1):20-5.

166. Caughey A, Sundaram V, Kaimal AJ, Gienger A, Cheng YW, McDonald K, Shaffer BL, Owens DK, Bravata DM. Systematic Review: Elective Induction of Labor Versus Expectant Management of Pregnancy. Annals of Internal Medicine. 2009;151(4):252-63.

167. Wood S, Cooper S, Ross S. Does induction of labour increase the risk of caesarean section? A systematic review and meta-analysis of trials in women with intact membranes. BJOG: Int J Obstet Gy. 2014;121(6):1471-0528.

168. Hussain AA, Yakoob MY, Imdad A, Bhutta ZA. Elective induction for pregnancies at or beyond 41 weeks of gestation and its impact on stillbirths: a systematic review with meta-analysis. BMC Public Health. 2011;11(3):1-12.

169. Boulvain, Marcoux, Bureau, Fortier, Fraser. Risks of induction of labour in uncomplicated term pregnancies. Paediatric and Perinatal Epidemiology. 2001;15(2):131-8.

170. Dublin S, Lydon-Rochelle M, Kaplan RC, Watts DH, Critchlow CW. Maternal and neonatal outcomes after induction of labor without an identified indication. Am J Obstet Gynecol. 2000;183(4):986-94.

171. Cammu H, Martens G, Ruyssinck G, Amy J-J. Outcome after elective labor induction in nulliparous women: A matched cohort study. AmJ Obstet Gynecol. 2002;186(2):240-4.

172. Glantz J. Term Labor Induction Compared With Expectant Management. Obstet Gynecol. 2010;115(1):70-6.

173. Grivell RM, Reilly AJ, Oakey H, Chan A, Dodd JM. Maternal and neonatal outcomes following induction of labor: a cohort study. Acta Obstet Gynecol Scand. 2012;91(2):198-203.

174. Danilack VA, Triche EW, Dore DD, Muri JH, Phipps MG, Savitz DA. Comparing expectant management and spontaneous labor approaches in studying the effect of labor induction on cesarean delivery. Annals of Epidemiology. 2016;26(6):405 - 11.e1.

175. Cheng YW, Kaimal AJ, Snowden JM, Nicholson JM, Caughey AB. Induction of labor compared to expectant management in low-risk women and associated perinatal outcomes. AmJ Obstet Gynecol. 2012;207(6):502.e1-.e8.

176. Rasmussen OB, Rasmussen S. Cesarean section after induction of labor compared with expectant management: no added risk from gestational week 39. Acta Obstet Gynecol Scand. 2011;90(8):857-62.

177. Stock SJ, Ferguson E, Duffy A, Ford I, Chalmers J, Norman JE. Outcomes of elective induction of labour compared with expectant management: population based study. BMJ. 2012;344.

178. Kruit H, Nuutila M, Rahkonen L. Synnytyksen käynnistäminen, kun raskaus on täysiaikainen. Suomen Lääkärilehti (Finnish Medical Journal). 2016;25-32:1845 - 51.

179. Martin JA, Hamilton BE, Osterman MJ, Curtin SC, Mathews TJ. Births: final data for 2013. Natl Vital Stat Rep. 2015;64(1):1-66.

180. Usha Kiran TS, Hemmadi S, Bethel J, Evans J. Outcome of pregnancy in a woman with an increased body mass index. BJOG: An International Journal of Obstetrics & Gynaecology. 2005;112(6):768-72.

181. Olesen AW, Westergaard JG, Olsen J. Prenatal risk indicators of a prolonged pregnancy. The Danish Birth Cohort 1998–2001. Acta Obstet Gynecol Scand. 2006;85.

182. Mogren I, Stenlund H, Högberg U. Recurrence of prolonged pregnancy. Int J Epidemiol. 1999;28:253-75.

183. Divon MY, Ferber A, Nisell H, Westgren M. Male gender predisposes to prolongation of pregnancy. AmJ Obstet Gynecol. 2002;187(4):1081-3.

184. Stotland NE, Washington AE, Caughey AB. Prepregnancy body mass index and the length of gestation at term. Am J Obstet Gynecol. 2007;197(4):378.e1-.e5.

185. Baranova A, Gowder SJ, Schlauch K, Elariny H, Collantes R, Afendy A, Ong JP, Goodman Z, Chandhoke V, Younossi ZM. Gene Expression of Leptin, Resistin, and Adiponectin in the White Adipose Tissue of Obese Patients with Non-Alcoholic Fatty Liver Disease and Insulin Resistance. Obesity Surgery. 2006;16(9):1118-25.

186. Laursen M, Bille C, Olesen AW, Hjelmborg J, Skytthe A, Christensen K. Genetic influence on prolonged gestation: A population-based Danish twin study. Am J Obstet Gynecol. 2004;190(2):489-94.

187. Hedegaard M, Lidegaard Ø, Skovlund CW, Mørch LS. Perinatal outcomes following an earlier post-term labour induction policy: a historical cohort study. BJOG: An International Journal of Obstetrics & Gynaecology. 2015;122(10):1377-85.

188. Mandujano A, Waters TP, Myers SA. The risk of fetal death: current concepts of best gestational age for delivery. American Journal of Obstetrics and Gynecology. 2013;208(3):207.e1-.e8.

189. Blickstein I, Flidel Rimon O. Post-Term Pregnancy. In: RJ M, AA F, MC W, editors. Neonatal-Perinatal Medicine: Diseases of the Fetus and Infant. 9th ed. St Louis: Elsevier Health Sciences; 2011. p. 351-7.

190. DiGiuseppe DL, Aron DC, Ranbom L, Harper DL, Rosenthal GE. Reliability of Birth Certificate Data: A Multi-Hospital Comparison to Medical Records Information. Maternal and Child Health Journal. 2002;6(3):169-79.

191. Verma A, Okun NB, Maguire TO, Mitchell BF. Morbidity Assessment Index for Newborns: A composite tool for measuring newborn health. American Journal of Obstetrics and Gynecology. 1999;181(3):701-8.

192. Novicoff WM, Wagner DP, Knaus WA, Kane EK, Cecere F, Draper E, Ferguson Ii JE. Initial development of a system-wide maternal-fetal outcomes assessment program. Am J Obstet Gynecol. 2000;183(2):291-300.

193. Löfling L, Bröms G, Bahmanyar S, Kieler H. Maternal and infant characteristics: differences and similarities between the Nordic countries and the US. Clinical Epidemiology. 2016;8:285-94.

194. Langhoff-Roos J, Krebs L, Klungsøyr K, Bjarnadottir RI, Källén K, Tapper A-M, Jakobsson M, Børdahl PE, Lindqvist PG, Gottvall K, Colmorn LB, Gissler M. The Nordic medical birth registers – a potential goldmine for clinical research. Acta Obstet Gynecol Scand. 2014;93(2):132-7.

195. Rosén M. National Health Data Registers: a Nordic heritage to public health. Scandinavian Journal of Social Medicine. 2002;30(2):81-5.

196. World Health Organization (WHO). International Classification of Diseases (ICD)-102010.

197. Government Decree on Screenings, 339/2011 (2011).

198. Ministry of Social Affairs and Health. Decree on the National Personal Records of Health Care Services 774/1989. 1989.

199. Gissler M. Registration of Births and Induced Abortions in the Nordic Countries. In: Kontula O, editor. Finnish Yearbook of Population Research XLV. Helsinki, Finland: The Population Research Institute of Finland; 2010.

200. Cnattingius S EA, Gunnarskog J, Källén B. A quality study of a medical birth registry. Scand J Soc Med. 1990;18(2):143-8.

201. Klungsøyr Melve K, Lie RT, Skjaerven R, Van Der Hagen CB, Gradek GA, Jonsrud C, Braathen GJ, Irgens LM. Registration of Down syndrome in the Medical Birth Registry of Norway: Validity and time trends. Acta Obstet Gynecol Scand. 2008;87(8):824-30.

202. Knudsen L, Olsen J. The Danish Medical Birth Registry. Dan Med Bull. 1998;45(3):320-3.

203. Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sørensen HT. The Nordic Countries as a Cohort for Pharmacoepidemiological Research. Basic & Clinical Pharmacology & Toxicology. 2010;106(2):86-94.

204. Delnord M, Szamotulska K, Hindori-Mohangoo AD, Blondel B, Macfarlane AJ, Dattani N, Barona C, Berrut S, Zile I, Wood R, Sakkeus L, Gissler M, Zeitlin J, and the Euro-Peristat Scientific C. Linking databases on perinatal health: a review of the literature and current practices in Europe. The European Journal of Public Health. 2016;26(3):422-30.

205. Johnson CE, Handberg E, Dobalian A, Gurol N, Pearson V. Improving perinatal and neonatal patient safety: The AHRQ patient safety indicators. Journal of Perinatal and Neonatal Nursing. 2005;19(1):15-23.

206. Kristensen S, Mainz J, Bartels P. Catalogue of Patient Safety Indicators - Safety Improvement for Patients in Europe. Report. Aarhus, Denmark: The ESQH-office for Quality Indicators; 2007.

207. World Health Organization (WHO). International Statistical Classification of Diseases and Related Health Problems 10th Revision. 2010.

208. American College of Obstetricians and Gynecologists (ACOG). Task Force on Cesarean Delivery. Evaluation of Cesarean Delivery. Washington, DC; 2000.

209. Kurth T WA, Glynn RJ, Chan KA, Gaziano JM, Berger K, et al. Results of Multivariable Logistic Regression, Propensity Matching, Propensity Adjustment, and Propensity-based Weighting under Conditions of Nonuniform Effect. American Journal of Epidemiology. 2006;163(3):262-70.

210. Imai K, Ratkovic M. Covariate balancing propensity score. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2014;76(1):243--63.

211. Schmidt C, Kohlmann T. When to use the odds ratio or the relative risk? International Journal of Public Health. 2008;53(3):165-7.

212. Team" RC. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2011.

213. Le Ray C, Blondel B, Prunet C, Khireddine I, Deneux-Tharaux C, Goffinet F. Stabilising the caesarean rate: which target population? BJOG: An International Journal of Obstetrics & Gynaecology. 2015;122(5):690--9.

214. Osterman MJ. Trends in Low-risk Cesarean Delivery in the United States, 1990–2013. Centers for Disease Control and Prevention (CDC),; 2014. Contract No.: 6.

215. Robson M, Murphy M, Byrne F. Quality assurance: The 10-Group Classification System (Robson classification), induction of labor, and cesarean delivery. Int J Gynaecol Obstet. 2015;131 Suppl 1:S23-7.

216. Tapia V, Betran AP, Gonzales GF. Caesarean Section in Peru: Analysis of Trends Using the Robson Classification System. PLoS One. 2016;11(2):e0148138.

217. Chong C, Su LL, Biswas A. Changing trends of cesarean section births by the Robson Ten Group Classification in a tertiary teaching hospital. Acta Obstet Gynecol Scand. 2012;91(12):1422-7.

218. Triunfo S, Ferrazzani S, Lanzone A, Scambia G. Identification of obstetric targets for reducing cesarean section rate using the Robson Ten Group Classification in a tertiary level hospital. European Journal of Obstetrics and Gynecology and Reproductive Biology. 2015;189:91-5.

219. Brennan DJ, Robson MS, Murphy M, O'Herlihy C. Comparative analysis of international cesarean delivery rates using 10-group classification identifies significant variation in spontaneous labor. AmJ Obstet Gynecol. 2009;201(3):308.e1-.e8.

220. Caughey AB, Cahill AG, Guise J-M, Rouse DJ. Safe prevention of the primary cesarean delivery. American Journal of Obstetrics and Gynecology. 2014;210(3):179-93.

221. Zhang J, Geerts C, Hukkelhoven C, Offerhaus P, Zwart J, De Jonge A. Caesarean section rates in subgroups of women and perinatal outcomes. BJOG: An International Journal of Obstetrics and Gynaecology. 2016;123(5):754-61.

222. Robson M, Hartigan L, Murphy M. Methods of achieving and maintaining an appropriate caesarean section rate. Best Pract Res Clin Obstet Gynaecol. 2013;27.

223. Howell S, Johnston T, Macleod SL. Trends and determinants of caesarean sections births in Queensland, 1997-2006. Aust N Z J Obstet Gynaecol. 2009;49(6):606-11.

224. Eglė B, Gintarė K, Regina RD, Arnoldas B, Laima M, Rūta N. Cesarean section rates in Lithuania using Robson Ten Group Classification System. IMedicina. 2015;51(5):280 - 5.

225. National Institute for Health and Welfare (THL). Perinatal statistics in the Nordic Countries 2014. 2016.

226. Gissler M. Arbetet med obstetriska indikatorer. Arbejdsgruppe under Nordisk Ministerråd om patientsikkerhed: En rapport om patientsikkerhed Styrelsen for Patientsikkerhed 2016; 2016.

227. Smith LA, Price N, Simonite V, Burns EE. Incidence of and risk factors for perineal trauma: a prospective observational study. BMC Pregnancy and Childbirth. 2013;13:59-.

228. Sheiner E, Levy A, Walfisch A, Hallak M, Mazor M. Third degree perineal tears in a university medical center where midline episiotomies are not performed. Archives of Gynecology and Obstetrics. 2005;271(4):307-10.

229. Ekéus C, Nilsson E, Gottvall K. Increasing incidence of anal sphincter tears among primiparas in Sweden: A population-based register study. Acta Obstet Gynecol Scand. 2008;87(5):564-73.

230. OECD. Health at a Glance 2015 - OECD Indicators: OECD Publishing; 2015.

231. Poulsen MØ, Madsen ML, Skriver-Møller A-C, Overgaard C. Does the Finnish intervention prevent obstetric anal sphincter injuries? A systematic review of the literature. BMJ Open. 2015;5(9):e008346.

232. Laine K, Gissler M, Pirhonen J. Changing incidence of anal sphincter tears in four Nordic countries through the last decades. Eur J Obstet Gynecol Reprod Biol. 2009;146(1):71-5.

233. De Leeuw JW, Struijk PC, Vierhout ME, Wallenburg HCS. Risk factors for third degree perineal ruptures during delivery. BJOG: An International Journal of Obstetrics & Gynaecology. 2001;108(4):383-7.

234. Aukee P, Sundström H, Kairaluoma MV. The role of mediolateral episiotomy during labour. Analysis of risk factors for obstetric anal sphincter tears. Acta Obstet Gynecol Scand. 2006;85(7):856-60.

235. Räisänen S, Vehviläinen-Julkunen K, Gissler M, Heinonen S. Hospital-based lateral episiotomy and obstetric anal sphincter injury rates: a retrospective population-based register study. American Journal of Obstetrics and Gynecology. 2012;206(4):347.e1-.e6.

236. Räisänen S, Cartwright R, Gissler M, Kramer MR, Laine K, Jouhki M-R, Heinonen S. Changing associations of episiotomy and anal sphincter injury across risk strata: results of a population-based register study in Finland 2004–2011. BMJ Open. 2013;3.

237. Finnström O, Berg G, Norman A, Olausson PO. Size of delivery unit and neonatal outcome in Sweden. A catchment area analysis. Acta Obstet Gynecol Scand. 2006;85(1):63-7.

238. Snowden JM, Cheng YW, Kontgis C, Caughey AB. The Association Between Hospital Obstetric Volume and Perinatal Outcomes in California. Am J Obstet Gynecol. 2012;207(6):478.e1-.e7.

239. Heller G, Misselwitz B, Schmidt S. Early neonatal mortality, asphyxia related deaths, and timing of low risk births in Hesse, Germany, 1990-8: observational study. BMJ : British Medical Journal. 2000;321(7256):274-5.

240. O'Donnell CPF, Kamlin COF, Davis PG, Carlin JB, Morley CJ. Interobserver variability of the 5-minute Apgar score. The Journal of pediatrics. 2006;149(4):486-9.

241. Siddiqui A, Cuttini M, Wood R, Velebil P, Delnord M, Zile I, Barros H, Gissler M, Hindori-Mohangoo AD, Blondel B, Zeitlin J, the Euro-Peristat Scientific C. Can the Apgar Score be Used for International Comparisons of Newborn Health? Paediatric and Perinatal Epidemiology. 2017;31(4):338-45.

242. Mollberg M, Hagberg H, Bager B, Lilja H, Ladfors L. High birthweight and shoulder dystocia: the strongest risk factors for obstetrical brachial plexus palsy in a Swedish population-based study. Acta Obstet Gynecol Scand. 2005;84(7):654-9.

243. Mehta SH, Blackwell SC, Bujold E, Sokol RJ. What factors are associated with neonatal injury following shoulder dystocia? J Perinatol. 2006;26(2):85-8.

244. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA. 2003;290(14):1868-74.

245. Paranjothy S, Frost C, Thomas J. How much variation in CS rates can be explained by case mix differences? BJOG: An International Journal of Obstetrics & Gynaecology. 2005;112(5):658-66.

246. Blomberg M. Maternal Body Mass Index and Risk of Obstetric Anal Sphincter Injury2014. 8 p.

247. Liu X, Du J, Wang G, Chen Z, Wang W, Xi Q. Effect of pre-pregnancy body mass index on adverse pregnancy outcome in north of China. Archives of Gynecology and Obstetrics. 2011;283(1):65-70.

248. Lloreda-García JM, Sevilla-Denia S, Rodríguez-Sánchez A, Muñoz-Martínez P, Díaz-Ruiz M. Resultados perinatales entre macrosomas hijos de madre diabética y macrosomas hijos de madre no diabética. Endocrinología y Nutrición. 2016;63(8):409-13.

249. Revicky V, Mukhopadhyay S, Morris EP, Nieto JJ. Induction of labour and the mode of delivery at term. Obstet Gynecol. 2011;31(4):304-6.

250. Ehrenthal D, Jiang X, Strobino D. Labor Induction and the Risk of a Cesarean Delivery Among Nulliparous Women at Term. . Obstetrics & Gynecology. 2010;116(1):35-42.

251. Jonsson M, Cnattingius S, Wikström A-K. Elective induction of labor and the risk of cesarean section in low-risk parous women: a cohort study. Acta Obstet Gynecol Scand. 2013;92(2):198-203.

252. Heffner L, Elkin E, Fretts R. Impact of Labor Induction, Gestational Age, and Maternal Age on Cesarean Delivery Rates. . 2003;102:287-93.

253. Davey M-A, King J. Caesarean section following induction of labour in uncomplicated first births- a population-based cross-sectional analysis of 42,950 births. BMC Pregnancy and Childbirth. 2016;16:92.

254. Danilack VA, Dore DD, Triche EW, Muri JH, Phipps MG, Savitz DA. The effect of labour induction on the risk of caesarean delivery: using propensity scores to control confounding by indication. BJOG: An International Journal of Obstetrics & Gynaecology. 2015;2016(123):1521-9.

255. Savitz DA, Terry JW, Dole N, Thorp JM, Siega-Riz AM, Herring AH. Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. Am J Obstet Gynecol. 2002;187.

256. World Health Organization (WHO). Neonatal and perinatal mortality: country, regional and global estimates. France; 2006.

257. Bruinsma A, Keulen J, Kortekaas J, van Dillen J, Oudijk M, Bakker J, van Kaam A, Bossuyt P, Vandenbussche F, van de Post J, Mol B-W, de Miranda E. 40: Induction of labor at 41 weeks or expectant management until 42 weeks - preliminary results of the INDEX trial. AmJ Obstet Gynecol. 2017;216(1):S27-S8.

258. Rosenstein MG, Cheng YW, Snowden JM, Nicholson JM, Caughey AB. Risk of Stillbirth and Infant Death Stratified by Gestational Age. Obstetrics and gynecology. 2012;120(1):76-82.

259. Altman M, Edstedt Bonamy A-K, Wikström A-K, Cnattingius S. Cause-specific infant mortality in a population-based Swedish study of term and post-term births: the contribution of gestational age and birth weight. BMJ Open. 2012;2.

260. Elden H, Hagberg H, Wessberg A, Sengpiel V, Herbst A, Bullarbo M, Bergh C, Bolin K, Malbasic S, Saltvedt S, Stephansson O, Wikström A-K, Ladfors L, Wennerholm U-B. Study protocol of SWEPIS a Swedish multicentre register based randomised controlled trial to compare induction of labour at 41 completed gestational weeks versus expectant management and induction at 42 completed gestational weeks. BMC Pregnancy and Childbirth. 2016;16(1):1-10.

261. Vrouenraets F, Roumen F, Md P, Dehing C, van den Akker E, Aarts M, Scheve E. Bishop Score and Risk of Cesarean Delivery After Induction of Labor in Nulliparous Women. Obstet Gynecol. 2005;105(4):690-7.

262. Pevzner L, Rayburn W, Rumney P, Wing D. Factors Predicting Successful Labor Induction With Dinoprostone and Misoprostol Vaginal Inserts. Obstet Gynecol. 2009;114(2, Part 1):261-7.

APPENDICES

Appendix 1: Data content of the Medical Birth Register

Personal data of mother
- Personal identity code
- Surname and forenames
- Profession
- Municipality of residence
- Nationality
- Marital status
- Cohabiting

Previous pregnancies and deliveries
- Previous pregnancies
- Previous deliveries

Present pregnancy and its monitoring
- Check-ups during pregnancy
- Date of first check-up visit
- Mother's weight and height before pregnancy
- Mother's smoking habits during pregnancy
- Risk factors and interventions relating to pregnancy
- Diseases during pregnancy (ICD-10 codes)
- Hospital care during pregnancy

Delivery
- Maternity hospital
- Place of birth
- Best estimate of gestational age at the time of delivery
- Onset of last period
- Duration of delivery
- Method of delivery
- Pain relief in labour
- Other procedures relating to delivery
- Diagnoses relating to pregnancy and delivery
- Mother's diagnoses during delivery (ICD-10 codes)

The infant
- Date of birth, control character of the personal identity code, time of birth
- Sex
- Infant born alive or dead
- Number of foetuses = number of infants born
- Letter indicating the order of birth in multiple pregnancy
- Weight at birth
- Length at birth
- Head circumference
- Apgar score at 1 minute and 5 minutes
- pH of umbilical blood

Data of the infant by the age of 7 days or at discharge
- Care interventions relating to the infant by the age of 7 days
- Infant's diagnoses by the age of 7 days
- Infant at the age of 7 days or at discharge from hospital
- Length of stay in hospital for mother

Data content of the data file Small Preterm Infants

Personal data of mother
- Personal identity code
- Surname and forenames

Personal data of infant
- Date of birth, control character of the personal identity code, time of birth
- Surname and forenames
- Sex
- Best estimate of gestational age at the time of delivery
- Weight at birth
- Length at birth

Head circumference
Number of fetuses = number of infants born
Letter indicating the order of birth in multiple pregnancy
Maternity hospital
Other basic data

Pregnancy

Mother's diseases and complications during the present pregnancy
mother's medication before delivery

Delivery

Rupture of amniotic membrane (water breaking)
Diastolic flow in umbilical artery
Apgar score at 1 minute, 5 minutes and 10 minutes
pH and BE of umbilical artery blood
pH and BE of umbilical vein blood
Method of delivery
Presentation at birth
Resuscitation procedures/treatment in delivery room

Treatment received by the infant up to 42 weeks' gestation

Breathing disorders
Breathing support
Surfactant treatment
Broncopulmonary dysplasia
Medication
infusion routes
Necrotising enterocolitis
Procedures and other treatment
Sepsis
Ultrasonography of the brain
Examinations of the fundus of the eye
Auditory examination
Magnetic resonance imaging (MRI) of the brain

Diagnoses for the infant up to 42 weeks' gestation

Diagnoses
Death diagnoses

The infant's situation at 42 weeks' gestation

The infant's situation when its age corresponds to 42 weeks' gestation (discharged, in hospital, dead)
Diet at discharge or at 42 weeks' gestation
Weight, length and head circumference at discharge or at 42 weeks' gestation

All hospitals where the infant has been treated up to 42 weeks' gestation

Hospitals and wards where the infant has been treated
Transferred to the next hospital
Where the infant has been transferred
Date, and the name and position of the person who has filled in the form

Regular data sources of the register

Maternity hospitals (maternity wards and neonatal wards)
Population Information System of the Population Register Centre
Statistics Finland, Population Statistics: Causes of Death